A self-consistent method for the analysis of lightning stroke data sets containing misclassified strokes: the variation of lightning over southern Africa

Michael David Grant

October 14, 2010

A thesis submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, October 2010

Declaration

I declare that this thesis is my own, unaided work, except where otherwise acknowledged. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

Signed this 14th day of October 2010

Michael David Grant

Abstract

Historically the low peak current portion of the positive cloud-to-ground lightning stroke data set is discarded since it is dominated by misclassified strokes. This thesis presents a self-consistent resolution to the problem of analysing lightning stroke data sets where misclassified strokes are present, without discarding sections of the data set. It is shown in this thesis that the misclassification problem is present in all the data sets, but is most prominent in the positive cloud-to-ground data set. The effect of truncating the positive cloud-to-ground lightning stroke data set from the South African Lightning Detection Network is that 43 % is discarded by truncating the data set below 10 kA, and 53 % is discarded if the data set is truncated below 15 kA. The statistical distribution of lightning stroke peak current over southern Africa is computed with the self-consistent method. A new measure of lightning activity is established that, in addition to activity, describes the energy of strokes. A previously undocumented inverse relationship between lightning stroke activity and peak current is presented in this thesis. The self-consistent method is extended to describe the diurnal variation of intracloud and cloud-to-ground parameters. The presence of positive cloud-to-ground lightning at the beginning and end of storms is verified from lightning detection network measurements. The new measure is applied to single storm days and single storms, and from this measure the charge distribution of the lightning producing clouds is inferred. To complement the diurnal and temporal variations, the unique orographic sensitivity of the various lightning polarity type combinations is presented over the extensive altitude of the Drakensberg mountain range. The self-consistent method presented in this thesis has direct application in meteorology; transmission line design and fault investigations; as well as improving risk analyses by providing the true distribution of lightning stroke peak currents.

Acknowledgements

Grateful thanks are extended to CBI-electric: low voltage and my colleagues who work there. An exceptional company who supported my academic goals on a profound level. Particularly, Paul Krüger, Wolfgang Weber, Chris Oliver, and Eldridge van Niekerk.

My research colleague, Andrew Dickson for his patience, support, insight and enthusiasm.

My colleagues at the University of the Witwatersrand, Johannesburg, South Africa. It is a great privilege to participate in the School of Electrical and Information Engineering. Particularly: Ian Jandrell, Ken Nixon, and Harry Fellows.

The South African Weather Service is graciously thanked for access to their data. Particularly Andrew van der Merwe and Karel de Waal.

The South African National Defence Force is thanked for use of their FAR-82 facility and support of the triggered lightning programme.

For funding and direct contributions to the research group: Eskom through their TESP programme, National Research Foundation for THRIP, Doug Kay and Lectro Tech Pty (Ltd), CBI-electric: low voltage for funding of the CBI-electric chair of lightning.

Steve Abbott and Dylan Morgan for some pretty crazy adventures hunting lightning.

To Mariana, thank you for an awesome team.

Mom and Dad.

And this grey spirit yearning in desire
To follow knowledge like a sinking star,
Beyond the utmost bound of human thought.
–Alfred Tennyson

Contents

List of Figures

List of Tables

List of Symbols and Operators

The principal symbols and operators used in this thesis are summarised below. The units are shown in square brackets and the first equation where the symbol is used is given.

g_n	Non-linear function of arrival time and sensor location for sensor n, *Equation (2.1)*
x_n	Longitude of sensor n, *Equation (2.1)*
y_n	Latitude of sensor n, *Equation (2.1)*
t_n	Arrival time at sensor n, *Equation (2.1)*
μ	Median of normal distribution, *Equation (4.1)*
σ	Variance of normal distribution, *Equation (4.1)*
θ_c	Cauchy distribution shape parameter, *Equation (4.3)*
I_c	Cauchy distribution peak parameter, *Equation (4.3)*
x_c	Cauchy distribution location parameter, *Equation (4.3)*
k_γ	Gamma distribution shape parameter, *Equation (4.4)*
θ_γ	Gamma distribution scale parameter, *Equation (4.4)*
I_γ	Gamma distribution peak parameter, *Equation (4.4)*
C_i	Cost of particle i, *Equation (4.5)*
$P_m[X]$	Measured probability distribution, *Equation (4.5)*
$P_f[X]$	Fitted probability distribution, *Equation (4.5)*

$w(n)$	Window function, *Equation (5.1)*
N_w	Number of points in window, *Equation (5.1)*
$w_a(n)$	Area normalised window of $w(n)$, *Equation (5.2)*
$i(t)$	Peak current magnitude at diurnal time t [kA], *Equation (5.3)*
$a(t)$	Activity at diurnal time t [strokes/min], *Equation (5.3)*
p	Peak for inverse relationship [kA^2], *Equation (5.3)*
o	Offset for inverse relationship [kA], *Equation (5.3)*
l	Grid length [m], *Equation (8.4)*
b	Grid breadth [m], *Equation (8.4)*
h	Slope height [m], *Equation (8.4)*
X	Random variable, *Equation (A.5)*
$E[X]$	Expectation operator on random variable X, *Equation (A.5)*
$M[X, k]$	Moment operator, k^{th} moment of X, *Equation (A.10)*
k	Order of moment, *Equation (A.10)*

List of Source Code Listings

Chapter 1

Introduction

The lightning phenomenon has been thoroughly investigated (Cooray (2003) and Rakov & Uman (2003)), and as new scientific methods to measure and observe the phenomenon are developed so new insights are gained. As technology for the detection of lightning strokes and the estimation of location and impulse parameters was developed, the spatial and temporal characteristics of lightning were investigated.

Increased lightning stroke counts are found in the central latitudes within the intertropical convergence zone (ITCZ) rather than extreme latitudes because globally lightning activity is enhanced in regions of high convection. The effect of the shape and relief of mountain ranges, termed orographic effects, also enhances lightning activity by increasing turbulence in the cloud mass. The Drakensberg mountain range is a severe and isolated mountain range found in the east of central South Africa, and the enclave of the mountainous Kingdom of Lesotho is also found in these mountains. This area presents a unique opportunity to study the variation of lightning with altitude as there are no adjacent orographic features to interfere with the orographic enhancement.

There are space based sensors that detect the optical transients associated with lightning activity, and this equipment has been used to categorise global lightning frequency and distribution (Christian *et al.* (2003) and Boccippio *et al.* (2001)). Christian *et al.* (2003) details how space based sensors have limited detection opportunities.

Lightning is classified by type and polarity; where polarity is defined as the direction along the normal vector from centre of the earth and type is either intracloud or cloud-to-ground. For cloud-to-ground discharges a direction is also used to describe the discharge; and thus the common downward negative stroke and rarer upward

positive stroke are assigned the same classification of negative cloud-to-ground by lightning detection networks (Rakov & Uman (2003)).

Intracloud lightning, where charge is transferred between regions in the atmosphere, produces a visible flash of light which is similar to the optical transient of cloud-to-ground lightning, where charge is transferred to, or from, the body of earth. Thus space based detectors are unable to discriminate between intracloud lightning and cloud-to-ground lightning (Christian *et al.* (2003)).

Terrestrial based lightning detection networks measure the radiated electromagnetic field from the ionisation of the channel and subsequent current impulse. While the detection range of terrestrial networks is limited by the locations of the sensors, their observation durations are continuous unlike the space based sensors. Boccippio *et al.* (2001) combine space and terrestrial detection systems to estimate the ratio of intracloud strokes to cloud-to-ground strokes.

Modern, commercial lightning detection networks, consist of a hybrid detection system where the radiated electric field is measured independently of the radiated magnetic field. There are several such networks in operation around the world, and one such network is located in South Africa. The South African Lightning Detection Network (SALDN) currently consists of 19 sensors owned and operated by the South African Weather Service. It is on data originating from this network that this thesis is based.

The hybrid detection system inherently has a misclassification problem, where small peak current positive intracloud strokes are incorrectly assigned the type classification of cloud-to-ground. This problem was first noticed after the lightning detection network in the United States of America was upgraded and a dramatic increase in positive cloud-to-ground strokes was observed by Cummins *et al.* (1998) and Wacker & Orville (1999*a*,*b*).

Cummins *et al.* (1998) and Cummins & Murphy (2009) advise other operators of the same technology to discard positive cloud-to-ground strokes based on a 10.0 kA limit of peak current estimated for the strokes. Wacker & Orville (1999*a*) described the marginal increase in negative cloud-to-ground strokes and ascribed it to improved detection efficiency and not misclassification as in the positive data set.

In line with international practise, climatological studies in South Africa have discarded positive strokes with peak currents less than the limit determined for the North American Lightning Detection Network (NALDN), for example the work by

Gill (2008). A subsequent video study, conducted over two storm seasons in two regions of North America evaluated the limit and increased it by 50 %. The extent and limit of the misclassification problem has not been assessed in the SALDN, and research continues by discarding positive strokes with peak currents below the first limit of 10.0 kA.

In this thesis the misclassification problem is quantified through a statistical approach where descriptive statistics—including the mean, median, and mode—are used to quantify the distribution of peak currents measured by lightning detection networks. These measures are used to separate the correctly classified component from the misclassified component in a data set. The diurnal, temporal and orographic variation of lightning strokes is then analysed with the self-consistent method.

1.1 Hypothesis and Contribution

It is proposed that the misclassification problem is present in all the data sets, regardless of polarity and network assigned type classification. The misclassification problem is resolved by a self-consistent method by which the independent variation of separate components within each data set may be determined. This thesis also shows that even though sections of the data set have been discarded through the application of the reclassification limits, previous statistical measures are still inherently biased by the misclassification problem.

The primary contribution and central theme in this thesis is unique, because unlike all other studies the low energy positive events (peak currents < 10.0 kA and subsequently < 15.0 kA) are not discarded from the data set. These low energy events are shown to provide a measure of intracloud processes vital in understanding the development of lightning producing clouds.

The contribution of this thesis is in

- presenting a consistent mathematical framework though which the independent variation of the components may be analysed;
- showing the distribution of cloud-to-ground and intracloud lightning as a combined probability distribution;

- establishing a new measure of lightning activity which accounts for both activity and energy;
- fitting two alternative statistical distributions to the random distribution of peak current of intracloud and cloud-to-ground lightning;
- verifying the two distributions against an independent video study;
- analysing the diurnal variation of lightning over southern Africa incorporating the large collection of data that was discarded previously; and
- applying the new measure to single storm days and a single storm.

The contribution has direct application in determining the true probability distribution of lightning strokes in a given area—a critical input to lightning risk analyses. Similarly the statistical distribution of lightning peak currents is used in many electro-geometric models to predict transmission line shielding failure and design appropriate protection (Cooray (2003)).

Peak lightning current distributions measured by lightning detection networks are inherently biased by the presence of misclassified strokes. The self-consistent method presented in this thesis provides a means to decompose the true peak current distribution, providing better quality inputs for engineering and scientific models.

1.2 Structure

The thesis begins with a brief description of the lightning phenomenon and detection technology, followed by the descriptive statistics and then the decomposed distributions. The diurnal and temporal variations are then presented, followed by the orographic variation of lightning stroke parameters through altitude of the Drakensberg mountain range. After the conclusion there are two appendices which provide additional detail on the statistical and algorithmic implementations.

Each argument of the thesis is presented in an individual chapter, as follows:

Chapter 2 - Charge, lightning and detection: This chapter presents a brief summary of the processes responsible for the lightning phenomena: charge separation, charge distribution, formation of the lightning channel and the transfer of

charge. The subsequent analysis in this thesis is performed on a set of lightning strokes detected over southern Africa. This chapter describes the detection technology installed in the Republic of South Africa, and the theory used to estimate lightning stroke location.

Chapter 3 - Statistical Parameters: This chapter begins with a description of the measures used to characterise the various distributions and then provides statistical descriptions of the data sets obtained by applying the network assigned type classification, as well as for the data sets obtained without applying the network assigned type classification. In addition to the descriptive statistics the contribution of this chapter is the calculation of the effect of truncating the positive cloud-to-ground data set, and a consistent mathematical framework that allows for the truncated section of the data set to be analysed *in situ*.

Chapter 4 - Decomposition of the data sets: This chapter proposes new distributions to describe intracloud and cloud-to-ground probability mass functions and presents an efficient numerical method for fitting these new distributions to the measured distributions. The contribution of this chapter is the decomposition of each of the data sets into separate intracloud and cloud-to-ground distributions, and shows that regardless of network assigned type classification each data set contains an intracloud and cloud-to-ground component. The fit quality of the complex distribution is compared to other distributions and the predicted misclassification ratios are compared to an independent video study.

Chapter 5 - Diurnal variations in the statistical parameters of lightning: This chapter begins by introducing the concept of a two parameter measurement, of activity and location, for the various statistical parameters. The temporal development in the diurnal cycle of the activity and location of the mean, median and mode for the various distributions arising from network assigned type classification is given. The contribution of this chapter is in showing the separate temporal development of the intracloud mode and cloud-to-ground mode, despite both originating from the same data set. The diurnal variation is developed into a general charge distribution model for lightning producing clouds in southern Africa.

Chapter 6 - Temporal development of the statistical mode for select days: In this chapter the mode is measured for select storm days and the diurnal variation presented in the general case is extended to specific days. The contribution of this chapter is in showing the applicability of the modal measurement on a limited data set such as a single day.

Chapter 7 - Temporal development of a select storm: This chapter presents a storm tracking algorithm that is used to select strokes that were part of a single storm. This subset of lightning strokes is then processed with the minute of the day algorithm introduced in previous chapters. The contribution of this chapter is in showing the spatial development of a storm and the corresponding lightning activity development. It is shown that the modal analysis method, developed in this thesis, is a valuable technique to be applied to single storms.

Chapter 8 - Orographic effects: The previous chapters focused on the temporal variation of lightning activity, and presented a novel method for measuring the variation of separate processes within the same data set. This chapter looks at the orographic variation of stroke parameters. The contribution of this chapter is in showing a specific altitude sensitivity in cloud-to-ground lightning stroke densities below 2150 m above sea level (m a.s.l.) and for negative intracloud stroke densities above 2 150 m a.s.l.

Chapter 9 - Conclusion: The distribution of lightning strokes is summarised in terms of the statistical descriptions. The summary includes the decomposition and comparison against the independent video study; the decomposition and modal analysis; and the orographic sensitivity of lightning strokes.

Appendix A - Statistical measures and distributions: The statistical measures and computational implementation used to characterise the probability distributions is presented in this appendix. The definitions of the mean, median, mode, moment, skewness, and kurtosis are repeated for completeness. Examples of these measures are given in terms of the widely used log-normal distribution for reference.

Appendix B - Algorithms and implementation: This appendix provides the implementation of the minute of the day analysis algorithm, storm phase algorithm and particle swarm optimisation algorithm.

Chapter 2

Charge, lightning and detection

This chapter presents a brief summary of the processes responsible for the lightning phenomena: charge separation, charge distribution, formation of the lightning channel and the transfer of charge. The subsequent analysis in this thesis is performed on a set of lightning strokes detected over southern Africa. This chapter describes the detection technology installed in the Republic of South Africa, and the theory used to estimate lightning stroke location.

2.1 Precursors to lightning

Lightning activity is dependent on the separation of charge; and within the extent of a cloud many mechanisms have been proposed, for example: inductive separation (Jennings (1975)), convective electrification (Chiu & Klett (1976)), and graupel ice collisions (Tsenova *et al.* (2009)). All the charge separation mechanisms are driven through the transfer of energy, the primary source of which is incident radiation from the sun. Solar energy warms the surface of the earth; and through thermal conduction, convection, and radiation this energy is transferred into the atmosphere. Initially, through adiabatic cooling (cooling of the gas through expansion without energy exchange with the environment) the air close to the surface of the earth expands and hence becomes less dense; eventually these bubbles of hot, light air become sufficiently buoyant to leave the surface.

The rate of rise of these bubbles is dependent on the pressure differential, and so for hotter surfaces and cooler climates these upliftings can become quite noticeable. They are also enhanced by meteorological events for example: cold fronts and low

pressure systems which enhance the flow rate and temperature differential. All of these parameters contribute to the movement of air, and a measure of this is the Convective Available Potential Energy (Donner & Phillips (2003)).

Many parameters are used to categorise lightning and fundamentally strokes are sorted into two types: cloud-to-ground lightning where charge is transferred from the atmosphere to the earth (or *vice versa*); and intracloud lightning where charge is transferred within the atmosphere. For these two categories, Rakov & Uman (2003) define the charge sign convention for cloud-to-ground strokes as follows: A positive sign indicates a movement of positive charge towards the centre of the earth, which is equivalent to the movement of negative charge away from the centre of the earth.

A negative sign indicates the opposite, where negative charge is moved towards the centre of the earth and positive charge away. For intracloud strokes the sign assigned is that, measured from the sign of the change in electric field, of the equivalent cloud-to-ground stroke (Rakov & Uman (2003)). Betz *et al.* (2008) notes that the accuracy of sign assigned to intracloud discharges depends on the extent of the vertical component of the discharge.

2.2 Charge separation processes and the role of convection

Hydrometeors consist of condensed or deposited atmospheric water vapour and form the primary constituent of clouds. Clouds are buoyant in the atmosphere (with majority gasses of Nitrogen, Oxygen, Argon and Carbon dioxide), additionally large airborne particles are also found within the cloud extent and may provide condensation opportunities to excess water vapour. The distribution of these components around the cloud is dependent on the particle mass, surface area and the temperature of the surrounding air.

It is through the interaction of the cloud components that charge is separated and a cloud becomes electrified. Various interaction mechanisms have been proposed to explain the separation phenomena, and many have results that compare favourably with thunder cloud measurements. However, individually the proposed mechanisms do not completely described the observed phenomena and it is probably through the cumulative action of the various mechanisms that the gross charge structure of a cloud forms.

Cooray (2003) describes many of the proposed theories, and for review three common theories are presented:

Inductive separation: The separation of charge through induction requires collisions between cloud particles in the presence of an electric field. The electric field induces a distribution of charge on the surface of the particle in opposition to the electric field. Collisions between particles allow for the momentary transfer of charge between the two particles as if it were a single particle in the field. The distribution of charge is schematically illustrated in *Figure 2.1.* However, an adequate reason preventing collision participants from coalescing is yet to be presented (Jennings (1975)).

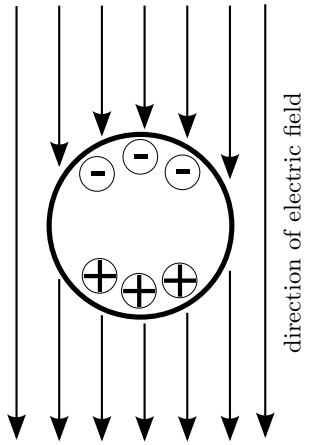
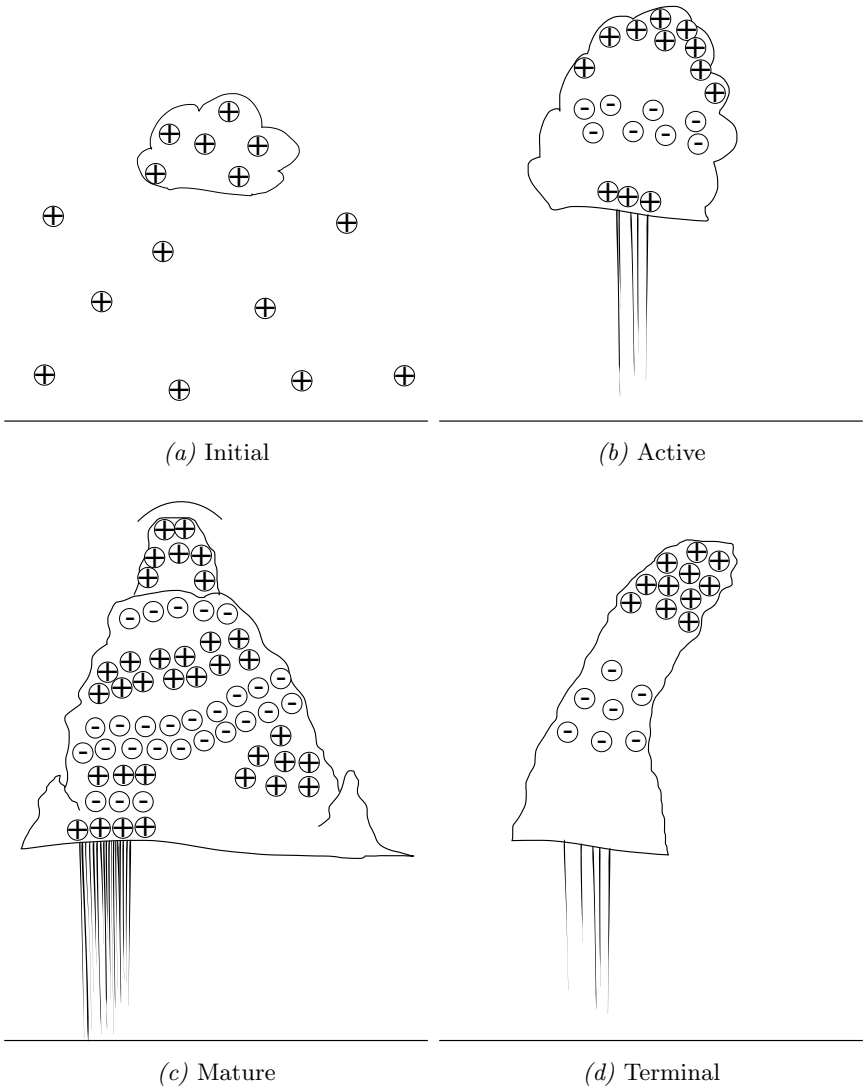


Figure 2.1: Schematic of charge distribution on a spherical water droplet in an electric field. Adapted from Cooray (2003).

Convective electrification: Charge is transported from sites of production due to electric field enhancement into the cloud through convection. As the charge within the cloud increases, more charge is produced on the ground due to the electric field enhancement. This positive feedback mechanism fails to account for the initial electrification, but may contribute some cloud charge (Chiu & Klett (1976)).

Graupel ice collisions: It has been proposed that a quasi-liquid layer on small hydrometeors may exchange charge during collisions. The physics of the thin layer of water molecules on the surface of a frozen hydrometeor determines both the magnitude and rate of charge separation (Tsenova *et al.* (2009)).



(a) Initial *(b)* Active

(c) Mature *(d)* Terminal

Figure 2.2: Charge distribution within various cloud stages.

Regardless of the mechanism responsible for creating two particles carrying opposite charge, convection distributes these particles throughout the cloud where the particle surface area to mass ratio ultimately describes where the particle is transported to.

The development of a lightning producing cloud, in terms of the charge distribution, is shown in *Figure 2.2*. Initially there is no difference between the cloud and fair weather charge, but through the action of the charge separation mechanisms a tripole is formed, with layers of positive charge at the top and bottom of the cloud and an extensive negative charge region in the centre. As discussed by Cooray (2003),

Rakov & Uman (2003) and Betz *et al.* (2008) as convection deepens additional charge layers are added, effectively screening upper charge layers from producing leaders (see below) that attach to ground. Thus the final stages of a lightning producing cloud, after the convective process has collapsed consists of a few layers of charge without many lower screening layers.

2.3 A summary of the lightning discharge

The many facets of the lightning discharge have been well documented: from the initial identification as an electrical phenomenon by Franklin (1753); to the various mechanisms responsible for the formation of the ionised channel (Schonland (1956), Cooray (2003) and Rakov & Uman (2003)). The parameters of electric and magnetic fields preceding, during, and after a stroke (Thottappillil *et al.* (1997), and Thottappillil *et al.* (1998)) have been described, as well as the various mechanisms of charge transfer in the channel (Cooray (2003) and Rakov & Uman (2003)).

An intracloud lightning discharge is the movement of charge within the cloud extent and results in the neutralisation of two (or more) charge centres within the cloud. Intracloud lightning is the most common form of the lightning discharge and is usually highly branched with short channel lengths. Peak currents, inferred from the magnitude of the electric field from the vertical component, of intracloud discharges are in the region of 2 kA to 10 kA. Cloud-to-ground discharges involves the transfer of charge between the cloud and ground and have longer channel lengths. Peak currents of cloud-to-ground discharges have been directly measured on tall structures and in rocket triggered-lightning and peak currents as high as 60 kA have been reliably measured (Jerauld *et al.* (2005) and Rakov & Uman (2003)).

The leader mechanism is responsible for the formation of ionised channels, be it between the cloud and ground or simply between charge centres in the cloud. The leader mechanism may be divided into two types: stepped and dart. The stepped leader is self propagating and is responsible for the formation of an ionised channel, whereas the dart leader follows an existing ionised channel (Mazur (2002)). In cloud-to-ground lightning, in addition to the polarity of return stroke current, flashes are described in terms of the stepped leader direction. Leaders most often originate from charge centres in a cloud, but may also start from the ground. Upward leaders are regularly initiated from tall structures (Mazur (2002)).

2.3.1 Stepped leader

The stepped leader propagates through air in discrete steps, with step lengths up to 200 m and average propagation velocities of 2×10^5 m/s (Rakov & Uman (2003)). The propagation of a stepped leader is driven by the charge in the streamer region infront of the leader. The pauses in the propagation of the stepped leader are due to the depletion of the leader charge, and propagation resumes as additional charge from the initial charge centre arrives at the leader head (Mazur (2002)). For negative leaders the arrival of additional charge at the leader head may result in branching; branching in positive leaders is unusual (Mazur (2002) and Rakov & Uman (2003)).

As the stepped leader approaches the surface of the earth, geometric electric field enhancement occurs and upward streamers are initiated from points with high induced electric fields. The upward streamers propagate towards the approaching leader, and may transition to leaders (Cooray (2003)). When the upward streamer or leader connects with the downward leader a continuous ionised channel between the cloud and ground has formed. It is said attachment has occurred and channel charge and the source charge drains to earth.

The result is a large electromagnetic pulse that radiates from the lightning channel (Thottappillil *et al.* (1997)). The peak current is a function of channel length (essentially how much charge is distributed along the path of the stepped leader), and the initial charge from the charge centre in the cloud. A return stroke may transfer some 5.2 C to ground; although as much as 24 C has been recorded.

2.3.2 Dart leader

One of the most well known features of a lightning flash is that the flash may consist of several subsequent strokes after the initial return stroke Cooray (2003). The frequency of occurrence of subsequent strokes is expressed in terms of mean flash multiplicity (Cooray (2003); Rakov & Uman (2003)). If there are numerous charge centres near the top of the ionised channel, dart leaders may follow the previously ionised path and transfer these charges to ground—resulting in subsequent strokes.

Since the dart leader follows an existing ionised channel it does not exhibit stepping or branching and the phenomenon has a much faster propagation velocity, around 1.1×10^7 m/s. There is also less channel charge and thus subsequent strokes have lower peak currents, and subsequent strokes transfer as much as 11 C, but 1.4 C is the average.

2.4 Lightning discharge location

Installation of the 19 sensors for the South African Lightning Detection Network (SALDN) began in 2005 and the network was fully commissioned by the beginning of 2007. The sensors in the SALDN use a combined Time of Arrival Technology (TOA) and Magnetic Direction Finding (MDF). The data in this thesis is from the SALDN after all the sensors were installed and before any changes were made to sensor location.

This section presents the theory of operation for the two technologies, for a review of other detection techniques the reader is directed towards the work of Rakov & Uman (2003).

2.4.1 Direction finding

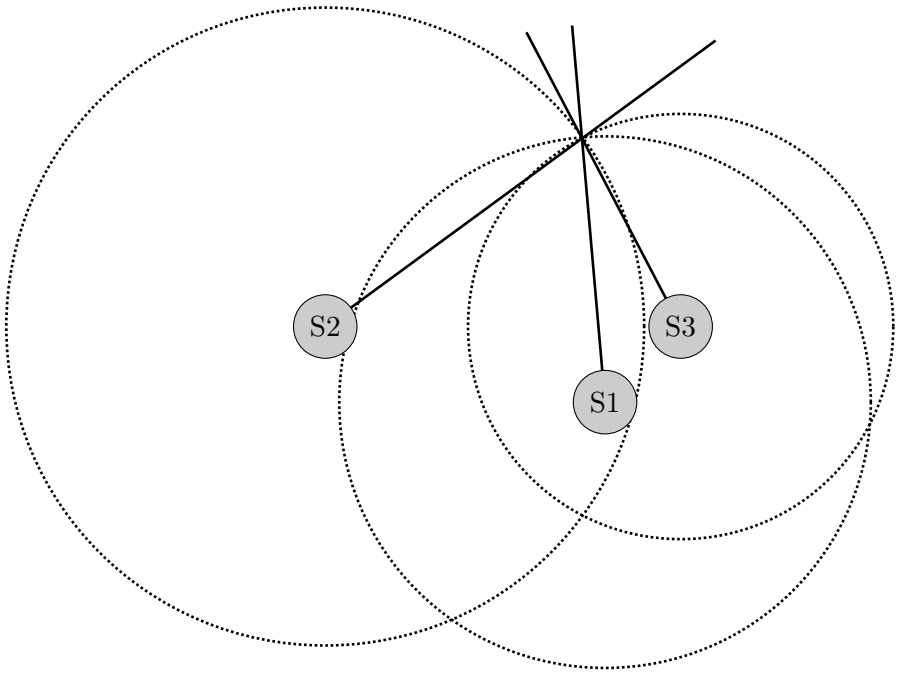


Figure 2.3: Schematic of system used to locate a lightning stroke; TDOA derived range circles are shown as dotted circles, and the thick black lines show bearings measured from magnetic direction finding.

Lightning location with magnetic direction finding is based on the assumption that the electric field component of the radiated electromagnetic wave is vertically orientated (with respect to the surface of the earth). The sensors typically consist of

two orthogonal loops vertically aligned connected to a gated wide band receiver as invented by Krider *et al.* (1976).

The gate operates on the assumption that the last few hundred meters of a cloud-to-ground stroke are approximately vertical and that the first few microseconds of the radiated impulse originate from this vertical section. Thus polarisation errors are minimised. The sign and direction of charge transfer ambiguity is eliminated through an additional measurement of the electric field component.

The location of the stroke is determined by the intersection of bearings from lightning detection stations, as shown in *Figure 2.3*. For the case of more than two sensors detecting the stroke the uncertainty in the measurement means that bearings converge on an area around a point, and so the stroke location is resolved as the minimisation of area problem.

2.4.2 Time of arrival

The time of arrival technique, or more precisely the time difference of arrival technique, describes the position of a radio source as a hyperbolic deviation from the baseline between the sensors. The detection system requires four sensors and solves for the location of the stroke in two dimensional space.

Koshak & Solakiewicz (2001) describe the derivation of the hyperbolic solution in a planar space as the solution to the system:

$$\begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} (x_2 - x_1) & (y_2 - y_1) & c^2(t_1 - t_2) \\ (x_3 - x_1) & (y_3 - y_1) & c^2(t_1 - t_3) \\ (x_4 - x_1) & (y_4 - y_1) & c^2(t_1 - t_4) \end{bmatrix} \begin{bmatrix} x \\ y \\ t \end{bmatrix} \tag{2.1}$$

where

g_n = Non-linear function of arrival time and sensor location for sensor n
x_n = Longitude of sensor n
y_n = Latitude of sensor n
t_n = Arrival time at sensor n

Where g are non-linear functions of arrival time and sensors located at (x_n, y_n) measuring the arrival time at t_n. Koshak & Solakiewicz (2001) also extend the solution to geodesically valid spaces.

Thompson *et al.* (1994) determines the time of arrival as the mean of the time at which the following occur:

1. on the rising edge: time to half peak;
2. time of peak; and
3. on the falling edge: time to half peak.

2.5 Lightning detection network observations

Cloud-to-ground strokes are discriminated from intracloud strokes based on rise time of the current impulse and peak magnitude. As the channel length for discharges confined to the atmosphere is usually short, the peak current is correspondingly lower. Intracloud stroke rise times are quite fast because the leader is often highly branched and propagation distances are small. Finally the orientation of the electromagnetic wave radiated from cloud to cloud strokes varies greatly as the nominal vertical component (with respect to the surface of the earth) may not even exist.

As detailed by Wacker & Orville (1999*b*), discrimination between small peak current cloud-to-ground discharges and intracloud discharges is problematic for the combined TOA/MDF technology. The problem has been described for the positive cloud-to-ground data set by Cummins *et al.* (1998), Wacker & Orville (1999*a*), Cummins *et al.* (2006), and Cummins & Murphy (2009).

In 1998 a threshold was set such that all positive strokes in the cloud-to-ground data set less than 10.0 kA were probably not cloud-to-ground and were incorrectly classified (Cummins *et al.* (1998)). In almost all subsequent studies based on TOA/MDF technology, the positive cloud-to-ground data set has been truncated by discarding all strokes less than 10.0 kA: for example the works of Boccippio *et al.* (2001), Lang & Rutledge (2002), Zajac & Rutledge (2001), and Gill (2008).

The threshold was subsequently revised in 2009 when a study by Biagi *et al.* (2007) based on video observations, suggested that the limit of misclassification was 15.0 kA.

No attempt has been made to assess extent of misclassification in the data set initially classified as intracloud. Cloud-to-ground strokes where the channel has long horizontal components, relatively small peak currents and fast rise times due

to topographic influences may in fact be detected by the network and classified as intracloud.

Summary

Energy from the sun drives convective processes which distribute kinetic energy throughout a cloud. In the process of energy distribution charge separation occurs and depending on the parameters of the convective process charge layers form within the extent of a lightning producing cloud. The same convective process is also responsible for the shape and vertical development of a storm cloud, and hence lightning activity is dependent on these parameters.

Lightning flashes, consisting of one or more strokes, occur as the rapid transfer of the separated charge between charge layers or between the cloud and ground. Lightning strokes radiate significant amounts of energy in both the visible and invisible electromagnetic spectrum. Lightning stroke locations are determined by analysing the lightning electromagnetic pulses measured by several spatially separated sensors.

The lightning detection network in the Republic of South Africa consists of 19 sensors each employing the combined TOA/MDF technology.

As documented by other international users of the same technology, there is a misclassification problem where positive intracloud strokes with small peak magnitudes are incorrectly classified as cloud-to-ground strokes. The majority of misclassified strokes is discarded from the positive cloud-to-ground data set through the application of a reclassification limit—below which all strokes are considered misclassified. The reclassification limit has recently been revised, and a bias is inherently present in previous work employing the old limit.

The following chapters present the statistical parameters of lightning observed in southern Africa with special attention to the misclassification problem. Subsequent chapters show that there are cloud-to-ground strokes below the reclassification threshold; that the misclassification problem is not just confined to small positive cloud-to-ground strokes and is present in the negative cloud-to-ground data set; and a consistent mathematical framework that allows for data sets to be analysed without applying inflexible and potentially inaccurate limits is presented.

Chapter 3

Statistical Parameters

This chapter begins with a description of the measures used to characterise the various distributions and then provides statistical descriptions of the data sets obtained by applying the network assigned type classification, as well as for the data sets obtained without applying the network assigned type classification. In addition to the descriptive statistics the contribution of this chapter is the calculation of the effect of truncating the positive cloud-to-ground data set, and a consistent mathematical framework that allows for the truncated section of the data set to be analysed *in situ*.

3.1 Statistical measures

Each lightning stroke recorded by the South African Lightning Detection Network (SALDN) is assigned a type classification of either intracloud or cloud-to-ground, and the peak current resolution is 2.0 kA and may range between −254.0 kA and +254.0 kA; although the response of lightning detection networks has only been verified up to 60.0 kA as discussed by Idone *et al.* (1993).

The study by Jerauld *et al.* (2005) on the combined TOA/MDF system was only able to evaluate strokes up to 45.0 kA. Data sets are created by selecting strokes based on peak current polarity and type classification with an additional case where the type classification is ignored.

A histogram is computed for a set of 2.0 kA bins distributed across the range of peak currents, and so for each data set the number of strokes occurring in each current bin is counted.

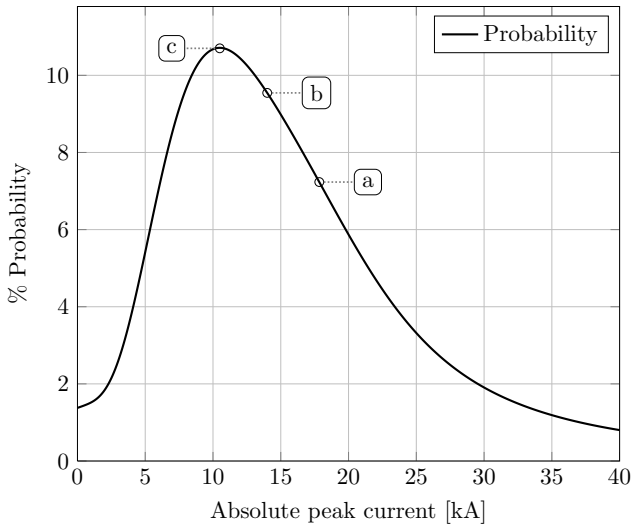


Figure 3.1: Illustration showing the location of the mean (a), median (b), and mode (c) on a positively skewed probability mass function. Note the order in which they appear on the positively skewed distribution.

The histogram is converted to a probability mass function by dividing each occurrence count by the total number of samples in each data set so that:

$$\sum_{x} P[X = x] = 1 \tag{3.1}$$

The mean, median, mode, standard deviation, skewness, and kurtosis is calculated for each data set. The implementation specifics are detailed in *Appendix A*.

Mean: The location of the mean on a probability distribution is shown in *Figure 3.1*, and is simply a measure of the average response of a random variable. For the descriptive statistics of the lightning stroke data sets, the mean is calculated as the expected value from the probability density function which is defined as the weighted sum of probabilities.

$$\mu = E[X] = \sum_{x} X \cdot P[X = x] \tag{3.2}$$

The calculation of the mean as the expected value will result in a value with a resolution finer than that of the peak current current measurement. This is a result of the geometric nature of the mean and should not be inferred as an increase of the precision of the measurement. Mean values in this thesis are reported with three significant digits so that the statistical distribution is reported accurately.

Median: The median is a location measure of a distribution and is defined as the point that divides the sample set in half such that there are equal probabilities of a sample lying above and below the median, and is shown in *Figure 3.1.* The median is a more robust measure of a distribution than the mean, and is less sensitive than the mean to the distribution variance, but is still influenced by the presence of extreme outliers.

The median has an inherent resolution limit of the histogram bin width and is thus reported without any significant digits. The median is calculated as the first bin where the cumulative probability is greater than or equal to 0.5:

$$\sum_{i=0}^{m} P[X = x_i] \geq \frac{1}{2} \tag{3.3}$$

Mode: The mode of a unimodal distribution is a measure of the most frequently occurring value, or the most probable outcome as shown in *Figure 3.1.* For symmetrical distributions this measure is the same as the mean and median. However, for asymmetrical distributions this measure is the peak of the probability density function. Similar to the median, the mode mode measured through inspection has an inherent resolution limit of the histogram bin width and is thus reported without any significant digits.

Complex distributions may have more than one mode, and in this case the modes of the distribution are defined as peaks distinct from the general trend and are numbered sequentially along the domain of the random variable. Modes are not numbered in terms of their probability thus the first mode may or may not occur more frequently than the second mode.

Standard deviation: The standard deviation of a distribution is a measure of the variability or dispersion about the mean. Larger values of the standard deviation indicate that the samples are far from the mean, while smaller numbers indicate that the samples are clustered around the mean.

$$\sigma = \sqrt{E[(X - \mu)^2]} \tag{3.4}$$

Skewness: The skewness of a random variable is a measure of the symmetry of the distribution, positive values of skewness indicate a right tailed distribution and negative values of skewness indicate a left tailed distribution. The skewness is calculated from the probability mass function as the third standardised moment:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}} \tag{3.5}$$

Kurtosis: The kurtosis of a random variable is a measure of the peakedness of the probability mass function and is defined as:

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \tag{3.6}$$

3.2 Misclassification

After the 1994/1995 combined technology upgrade of the lightning detection network in the United States of America a dramatic increase in the number of small peak current positive cloud-to-ground strokes was observed (Cummins *et al.* (1998), and Wacker & Orville (1999*a*,*b*)). A similar change was not observed in the negative cloud-to-ground set and the effect in the intracloud set was not documented as the network was specifically focused on the detection of cloud-to-ground strokes and the detection of intracloud strokes was regarded as incidental.

In order to achieve continuity between the pre- and post- upgrade data sets the set of positive cloud-to-ground strokes below 10.0 kA was reclassified as intracloud—this was implemented by many researchers through discarding this section of the data set (for example Lang & Rutledge (2002) and Gill (2008)). A subsequent study by Cummins & Murphy (2009) suggests that this limit should be increased to 15.0 kA. The effect of applying the 1998 limit of 10.0 kA and the 2009 limit of 15.0 kA is discussed in the following section for the positive cloud-to-ground data set, and is shown in the subsequent probability mass functions.

Positive cloud-to-ground lightning strokes are categorised by their significantly larger peak currents and are documented as the cause for some of the most severe lightning

damage (Cooray (2003) and Rakov & Uman (2003)). Negative cloud-to-ground discharges are much more common than positive cloud-to-ground discharges and subsequent strokes with small peak currents have been measured (Chisholm & Cummins (2006) and Rakov & Uman (2003)).

Intracloud lightning channel lengths are shorter than their cloud-to-ground counterparts and although peak currents are only estimated from the peak electric field of the vertical component of the discharge, the measured peak electric field is much less than cloud-to-ground strokes.

Thus the intracloud distributions are expected to have small medians and relatively small skewness, the negative cloud-to-ground distribution is expected to have a smaller median than the positive cloud-to-ground distribution, but a larger value of skewness than the intracloud distributions. The positive cloud-to-ground distribution is expected to have the largest median as well as a large value of skewness.

Since the presence of misclassified strokes in the positive cloud-to-ground data set has been verified and the intracloud and cloud-to-ground distributions are so different, it is expected that the resulting distribution will be bimodal (one mode for the intracloud component and another for the cloud-to-ground component). However since the negative cloud-to-ground strokes may have relatively small peak currents a clear delineation between the intracloud and cloud-to-ground components in the resulting negative cloud-to-ground data set is not expected.

3.3 Intracloud data set

For the probability mass functions from data set of intracloud assigned network type classification (negative and positive peak currents), the statistical parameters are detailed in *Table 3.1.* The probability mass functions of the intracloud data set are plotted in *Figure 3.2* with the 1998 and 2009 limits shown for reference. Both distributions are unimodal although the magnitude of the negative mode is larger than the magnitude of the positive mode.

The means and medians of the intracloud data sets are comparable, as is the skewness and kurtosis. Of the 4.8 million strokes assigned an intracloud type classification, strokes with a positive polarity comprise of 70 % of the data set: 2.35 times as many strokes assigned a positive polarity than those assigned a negative polarity.

Table 3.1: Statistical parameters for negative and positive strokes from the set of strokes classified as intracloud strokes, recorded by the SALDN.

	Negative	Positive
Samples	1,452,777	3,420,372
% of total strokes	2.24	5.27
Mean [kA]	-11.717	12.018
Median [kA]	-10	8
1st Mode [kA]	-8	6
2nd Mode [kA]	–	–
Standard deviation	7.921	10.715
Skewness	2.9581	2.8656
Kurtosis	23.530	16.014

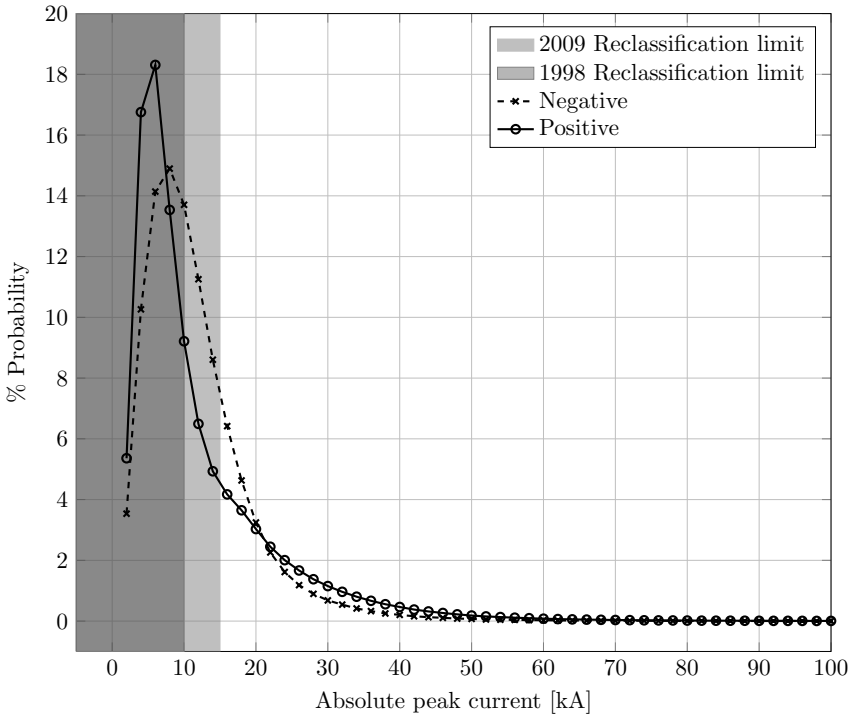


Figure 3.2: Probability mass functions for negative and positive strokes, from the set classified as intracloud strokes, recorded by the SALDN.

The positive intracloud probability mass function shown in *Figure 3.2* exhibits a discontinuity between 10.0 kA and 20.0 kA. The discontinuity arises where the rate of decreasing probability changes abruptly. No similar abrupt change is visible in the negative probability distribution, shown in *Figure 3.2*, although the probability of strokes with peak currents above 20.0 kA rapidly decays.

For both the negative and positive probability density functions the means are less than the 2009 reclassification limit of 15.0 kA. The median of the negative intracloud distribution occurs on the 1998 limit of 10.0 kA, and the positive median is less than the 1998 reclassification limit of 10 kA. Both the positive and negative statistical are less than the 1998 reclassification limit of 10.0 kA. In terms of the probability mass functions for the intracloud data set the reclassification limits describe at least 50 % of all the strokes.

3.4 Cloud-to-ground data set

The negative cloud-to-ground data set, with more than 52 million strokes, is 7.3 times larger than the positive cloud-to-ground data set with 7 million strokes. The negative cloud-to-ground probability mass function, shown in *Figure 3.3* is unimodal with positive skewness. The magnitudes of mean, median, and mode are 6.1 kA, 4 kA and 2.0 kA larger than the respective measures of the intracloud distribution.

The positive cloud-to-ground probability mass function, also shown in *Figure 3.3*, is bimodal with the first statistical mode at 6.0 kA and the second at 18.0 kA. Although the positive intracloud distribution is unimodal, the location of the first mode in the positive cloud-to-ground distribution is the same as the location of the mode in the positive intracloud distribution.

The application of the reclassification limits discussed previously in *Section 3.2* is advocated for the positive cloud-to-ground data set. As shown in *Table 3.2* a significant number of positive cloud-to-ground strokes are discarded by the reclassification process. Discarding strokes with peak currents less than the 1998 limit of 10.0 kA reduces the data set by 3.1 million strokes which is 43 % of the data set; this increases the mean by 8.0 kA to 24.7 kA. If the 2009 limit of 15.0 kA is applied, 3.9 million strokes are discarded which is 54 % of the data set; this action increases the mean by 11.0 kA to 27.7 kA.

Similarly in the application of the first limit the median increases by 6.0 kA to 20.0 kA

Table 3.2: Statistical parameters for negative and positive strokes from the set of strokes classified as cloud-to-ground, recorded by the SALDN.

	Negative	Positive	Positive >10.0 kA	Positive >15.0 kA
Samples	52,712,872	7,237,474	4,104,591	3,276,973
% of total strokes	81.31	11.16	6.33	5.05
Mean [kA]	-17.841	16.739	24.784	27.763
Median [kA]	-14	14	20	22
1st Mode [kA]	-10	6	18	18
2nd Mode [kA]	–	18	–	–
Standard deviation	13.289	15.123	15.788	16.369
Skewness	3.2772	3.3071	3.7264	3.7323
Kurtosis	21.753	20.688	22.738	21.910

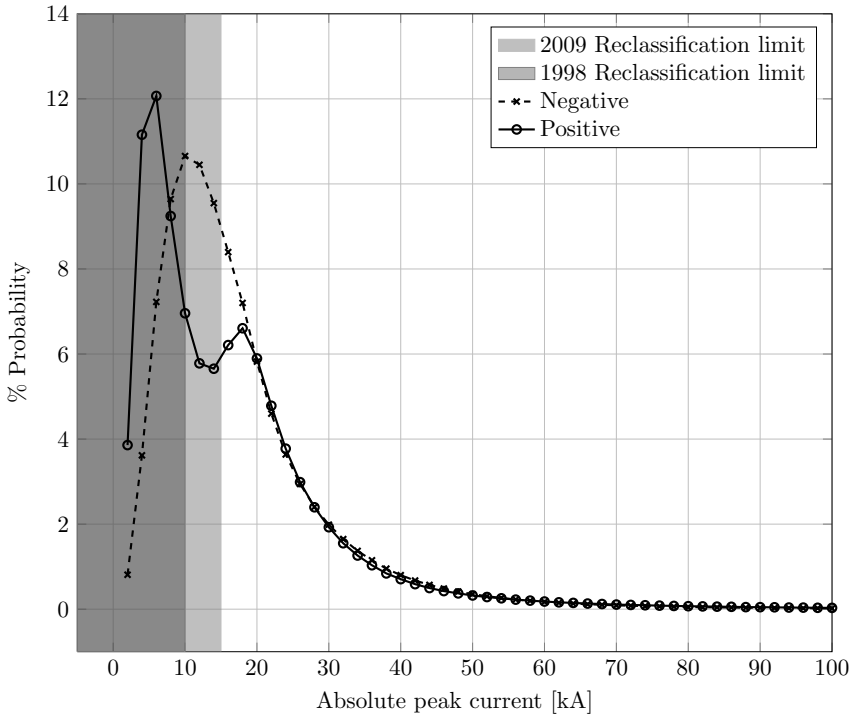


Figure 3.3: Probability mass functions for negative and positive strokes, from the set classified as cloud-to-ground, recorded by the SALDN.

and for the second limit the median increases by 8.0 kA to 22.0 kA. In addition to meteorological parameters, the commonly used statistical measures are inherently sensitive to the reclassification limits.

Since the location of second mode of the positive cloud-to-ground distribution is larger than both the 10.0 kA and 15.0 kA limits it is unchanged by the reclassification operation, and is simply renumbered in the absence of the first mode.

The shape of the tails of the negative and positive distributions are similar, with comparable rates of decreasing probabilities. The abrupt change in the positive intracloud distribution, shown in *Figure 3.2*, is not present in the negative or positive cloud-to-ground distribution, but the location of the abrupt change in the intracloud data set is comparable with the second mode in the cloud-to-ground data set.

3.5 Entire data set

If the network assigned type classification is ignored, then the distributions shown in *Figure 3.4* and detailed in *Table 3.3* are obtained.

Despite the locations of the modes for the negative intracloud and cloud-to-ground distributions differing by 2.0 kA, the peak current resolution and distribution kurtosis is insufficient to discern two separate peaks.

Since there are 36 times more strokes classified as negative cloud-to-ground, the mean, median and mode for the negative distribution, created by ignoring the network assigned type classification, are more representative of the cloud-to-ground distribution. Hence despite sampling two separate processes (intracloud and cloud-to-ground), this fact is not obvious through the inspection of the negative distribution.

In the complete positive data set there are only twice as many strokes with a cloud-to-ground network assigned type classification than strokes with an intracloud network assigned type classification. Thus the resulting positive statistical distribution describes an average response to positive strokes with a small bias to the cloud-to-ground type. Additionally the difference between the two modes is 12.0 kA, much larger than any aliasing effects and so both modes are visible in *Figure 3.4*.

Table 3.3: Statistical parameters for negative and positive strokes from the set of all strokes, regardless of type classification, recorded by the SALDN.

	Negative	Positive
Samples	54,165,650	10,657,846
% of total strokes	83.55	16.44
Mean [kA]	-17.677	15.224
Median [kA]	-14	12
1st Mode [kA]	-10	6
2nd Mode [kA]	–	18
Standard deviation	13.211	14.035
Skewness	3.2846	3.3423
Kurtosis	21.902	21.705

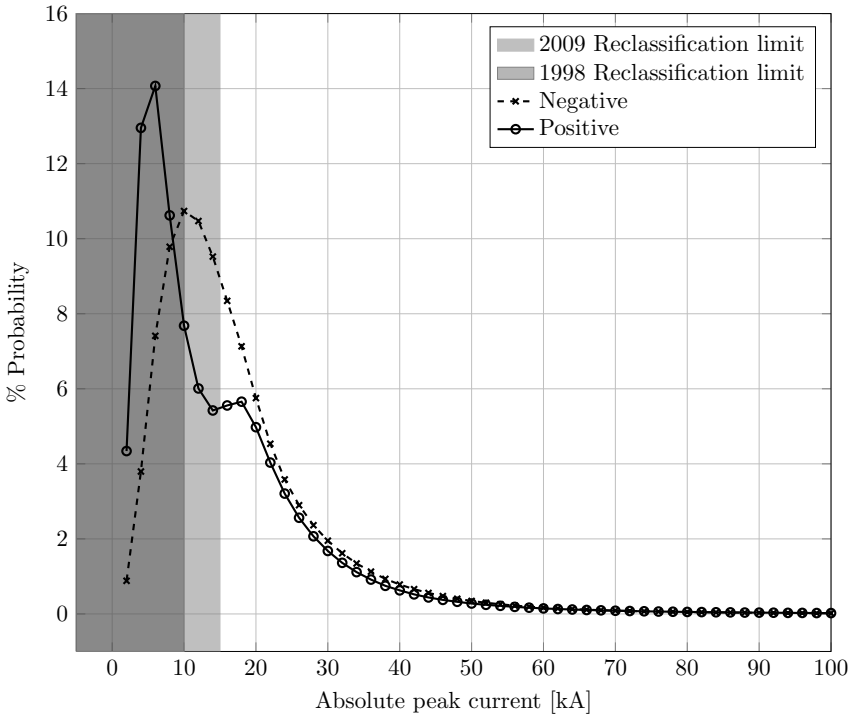


Figure 3.4: Probability mass functions for negative and positive strokes from the set of all strokes, regardless of type classification, recorded by the SALDN.

Summary

In this chapter the descriptive statistics for the distributions arising from the type assigned classifications were presented. These measures are used to develop climatology models and risk profiles, but because of a misclassification problem in the positive cloud-to-ground data set these values must be handled accordingly.

The negative cloud-to-ground distribution is similar to the negative intracloud distribution since both are unimodal with comparable means and medians.

Unlike the negative distributions the positive cloud-to-ground data set is bimodal, with the first mode describing the intracloud process and the second mode describing the cloud-to-ground process.

The effect of applying the reclassification limits to the statistical measures of the positive cloud-to-ground data set was shown to increase the mean and median, but the location of the second mode is not sensitive to the application of these reclassification limits.

In the analysis of a complex distribution, the mode provides a mechanism for measuring the parameters of the independent processes without artificially truncating or splitting the data set. If the limits for selecting elements of one process are set such that under certain conditions significant numbers of the population are excluded then the subsequent measures are flawed. The mode allows for the dynamic measurement of each process separately provided that each process is substantially accounted for by members of that sample set.

The following chapter extends the descriptions of the distributions by showing that each distribution consists of a combination of an intracloud distribution and a cloud-to-ground distribution. The decomposition of the distributions is achieved through the application of an efficient numerical method, and even the apparently unimodal distributions are shown to be a combination.

Chapter 4

Decomposition of the data sets

This chapter proposes new distributions to describe intracloud and cloud-to-ground probability mass functions and presents an efficient numerical method for fitting these new distributions to the measured distributions. The contribution of this chapter is the decomposition of each of the data sets into separate intracloud and cloud-to-ground distributions, and shows that regardless of network assigned type classification each data set contains an intracloud and cloud-to-ground component. The fit quality of the complex distribution is compared to other distributions and the predicted misclassification ratios are compared to an independent video study.

4.1 Probability mass functions

The random distribution of lightning stroke peak current is usually described by a cumulative log-normal probability density function, where the probability is of the stroke peak current exceeding a specific magnitude. The knee point in a log-log plot of the cumulative density function corresponds to the mode in a normal plot of the probability density.

Numerous designs are based on these peak current probability distributions and for consistency the parameters of the distribution have been standardised. For example the lightning and insulator subcommittee of the IEEE transmission and distribution committee defines parameters for the log-normal peak current distribution (Chowdhuri *et al.* (2005) and Chisholm & Cummins (2006)) where the distribution has a median of 31.1 kA and log standard deviation of $\sigma = 0.48$.

This section also details the Cauchy and gamma distributions. Similar to the log-normal distribution, these distributions are also naturally occurring with mathematical properties particularly suited to their application in this case.

4.1.1 Log-normal distribution

A log-normal distribution is obtained by taking the logarithm of a normal distribution with median μ and standard deviation σ. A peak scaling term, I is commonly introduced. The properties of the log-normal distribution are detailed in *Table 4.1*.

Log-normal distributions are typically positively skewed (right tailed) and are often fitted to natural phenomena that are positively skewed. The log-normal distribution is favoured because of the logarithmic relationship to a normal distribution - an indication of some process that is normally distributed.

$$f(x;\mu,\sigma) = \frac{I}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \tag{4.1}$$

where

μ = Median of normal distribution

σ = Variance of normal distribution

Table 4.1: Properties of the log-normal distribution.

Mean	$e^{\mu+\sigma^2/2}$
Median	e^{μ}
Mode	$e^{\mu-\sigma^2}$
Variance	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$
Skewness	$(e^{\sigma^2}+2)\sqrt{e^{\sigma^2}-1}$
Kurtosis	$e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-3$

4.1.2 Cauchy distribution

The Cauchy distribution is also known as the Lorentz distribution, and is characterised by three parameters: x_c which is the location parameter that describes the median and mode; θ which is the scale parameter that describes the half-width at half-peak; and I which is the peak parameter that scales the whole distribution. The properties of the Cauchy distribution are detailed in *Table 4.2*.

The moments of the distribution are undefined because the integral over the distribution evaluates to infinity; there is no direct physical interpretation for this, although in the context of peak lightning current distribution exceptionally high peak currents have been measured and this distribution does not place a bound on this upper limit. Thus all measures derived from the expected value in an unbounded domain are undefined (skewness and kurtosis undefined).

$$f(x; x_c, \theta_c, I_c) = \frac{I_c}{\pi\theta_c \left[1 + \left(\frac{x-x_c}{\theta_c}\right)^2\right]} \tag{4.2}$$

$$= \frac{I_c}{\pi}\left[\frac{\theta}{(x-x_c)^2 + \theta_c^2}\right] \tag{4.3}$$

where

θ_c = Cauchy distribution shape parameter
I_c = Cauchy distribution peak parameter
x_c = Cauchy distribution location parameter

Table 4.2: Properties of the Cauchy distribution.

Mean	Undefined
Median	x_c
Mode	x_c
Variance	Undefined
Skewness	Undefined
Kurtosis	Undefined

4.1.3 Gamma distribution

The gamma distribution is also a three parameter distribution with a scale parameter θ, shape parameter k_γ, and peak parameter I_γ. The properties of the distribution are detailed in *Table 4.3*.

The gamma distribution arises naturally from modelling the waiting time between random events for particular distributions (for example the time between two successive and equal rolls of a dice). In this analysis the gamma distribution is used to model the intracloud process because of the small variance and exponentially decaying tail.

$$f(x; k_\gamma, \theta_\gamma, I_\gamma) = I_\gamma x^{k_\gamma - 1} \frac{e^{-x/\theta_\gamma}}{\theta_\gamma^{k_\gamma} \Gamma(k_\gamma)} \quad \forall (x, k_\gamma, \theta_\gamma) > (0, 0, 0) \tag{4.4}$$

where

k_γ = Gamma distribution shape parameter
θ_γ = Gamma distribution scale parameter
I_γ = Gamma distribution peak parameter

Table 4.3: Properties of the gamma distribution.

Mean	$k_\gamma \theta_\gamma$
Median	Undefined
Mode	$(k_\gamma - 1)\theta_\gamma$
Variance	$k_\gamma \theta_\gamma^2$
Skewness	$\frac{2}{\sqrt{k_\gamma}}$
Kurtosis	$\frac{6}{k_\gamma}$

4.1.4 Poisson process

A Poisson process is a stochastic process in which events occur continuously and independently from one another. This is an appropriate description for strokes detected by a lightning detection network as many storms co-exist and intracloud lightning is detected independently of cloud-to-ground lightning.

The expectation value of a Poisson process, which is created by combining two processes (P_1 and P_2), is expressed as $E[P_1 + P_2]$ as is categorised by the sum of the two distributions. The particle swarm optimisation, detailed below, is applied to decompose the single distributions into the independent intracloud and cloud-to-ground components.

4.2 Particle swarm optimisation

For two parameter distributions it is possible to solve for the parameters using algebraic methods, and for well defined distributions it may be possible to solve for a third parameter. However, for complex distributions where the combined distribution is described by six parameters, it is not possible to solve for the parameters

algebraically. Instead a computational approach, where the quality of fit is described as how closely the combined distribution matches the measured distribution is employed.

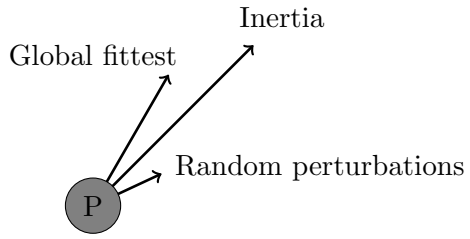


Figure 4.1: Illustration of velocity components and their relative scales, used to compute the location of a particle in the swarm.

Particle swarms perform very efficient computational searches of multidimensional problem spaces, and each particle needs to be aware of the location of the particle with the best cost in the swarm. Like real particles, each has mass and moves at varying velocities, depending on the resultant force applied - this is shown in *Figure 4.1.*

Particle swarm searches are more computationally efficient than random searches, but can also incorporate elements of the random search in order to improve the search efficiency further: initially the particles are randomly distributed around the search space, and then at each time step a random perturbation is applied. The small random perturbation is applied at each time step so that the local area around each particle is explored (instead of just the route towards the global fittest point). Additionally, particle swarm searches are relatively insensitive to local maxima and minima, and the introduction of the small random perturbations also improves this performance further.

$$C_i = \sum_x |P_m[X = x] - P_f[X = x]| \tag{4.5}$$

where

C_i = Cost of particle i

$P_m[X]$ = Measured probability distribution

$P_f[X]$ = Fitted probability distribution

Equation 4.5 defines the cost function for each particle as the magnitude of the difference in areas between the original distribution and the fitted distribution; the measured distribution is interpolated to a finer resolution since the 2.0 kA resolution

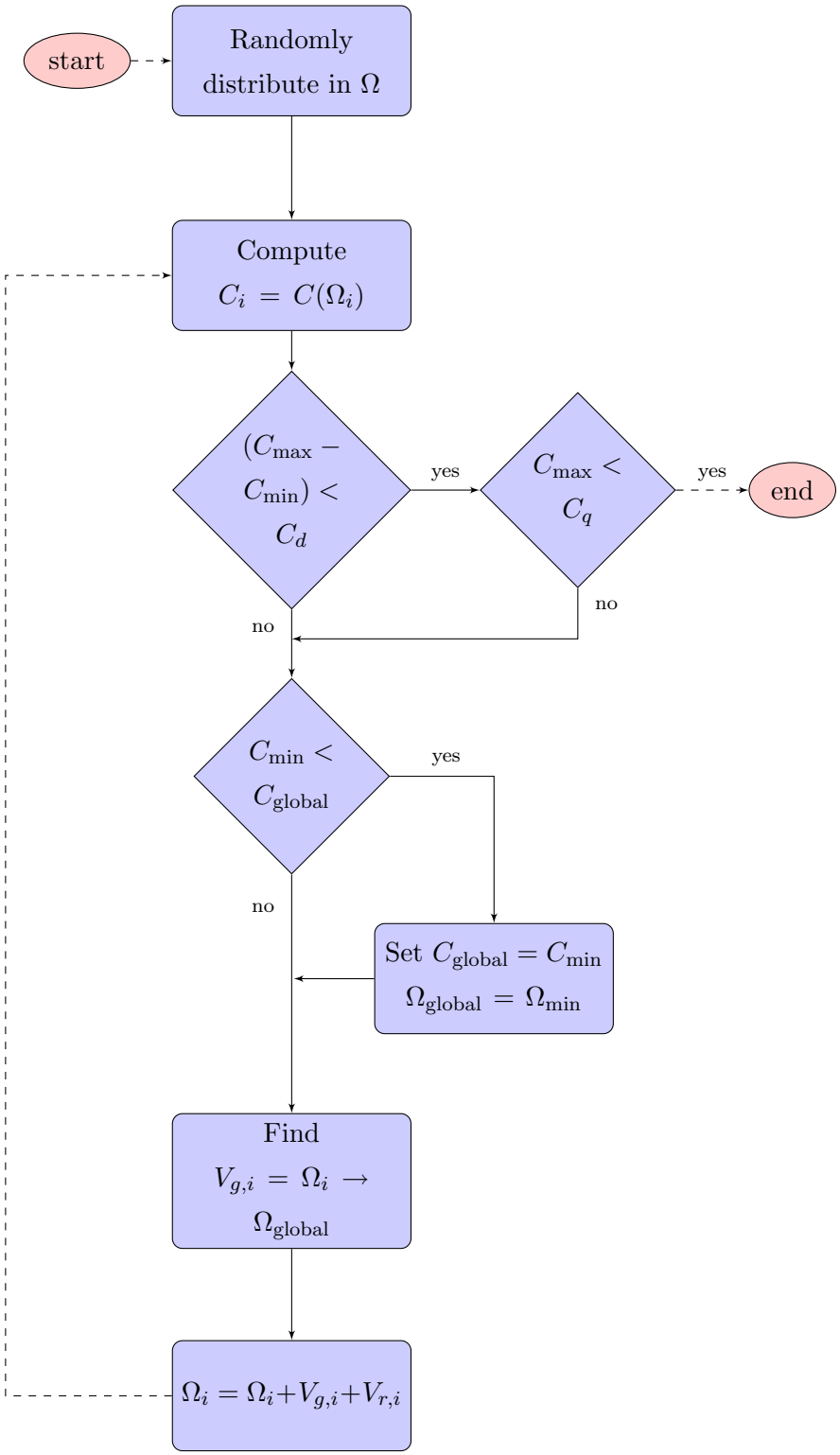


Figure 4.2: Particle swarm optimisation algorithm for a set of particles that start randomly distributed through the problem space.

is is imposed by the resolution of the measurement and not quantisation of the physical process. By computing the cost function across a finer resolution the fit quality is improved further, and potential discontinuities are avoided.

As shown in *Figure 4.2*, the algorithm begins by randomly distributing the particles in the problem space Ω, and then the cost function, detailed in *Equation 4.5*, is computed for each particle where the location in the problem space serves as the parameters for the cost function.

Then two swarm conditions are evaluated: the first condition is the convergence in the cost domain where, if the difference between the maximum cost and minimum cost is less than the parameter C_d, the swarm has converged. The second condition is the evaluation of the cost of the unfittest member of the swarm against the minimum fit quality parameter C_q, which sets the acceptable limit of the solution quality.

If the convergence and quality criteria are not met then the cost of fittest member of the swarm is evaluated against the global fittest (for all iterations) and if the fittest member of the swarm has improved on the global fittest, then both the global cost and location are set to the fittest member. So as the search progresses through subsequent iterations the global fittest is a measure of search performance and a record of the best found solution.

The next step is that a vector for each particle towards the global fittest is computed, and then a force comprising of a component towards the global fittest ($V_{g,i}$) and a random perturbation ($V_{r,i}$) is applied to the particle.

The source code for the implementation of this particle swarm search is in *Appendix B*.

4.3 Quality of fit

A single log-normal distribution, combined double log-normal distribution, and combined Cauchy-gamma distribution are fitted to the measured distributions. In the combined Cauchy-gamma distribution the Cauchy distribution represents the cloud-to-ground component and the gamma distribution represents the intracloud component.

The quality of fit is measured as the integral of the magnitude of the difference between the fitted distribution and the measured distribution. The quality of fit is

Table 4.4: Quality of fit for the Cauchy-gamma, historical log-normal and double log-normal distributions. Numbers in parentheses indicate percentage of best fit, which is consistently achieved by the Cauchy-gamma distribution.

Distribution	Cauchy-gamma	Log-normal	Dbl. log-normal
Neg. intracloud	0.28603 (–)	1.18025 (412%)	0.33023 (115%)
Neg. cloud-to-ground	0.35662 (–)	0.67138 (188%)	0.44133 (123%)
Neg. all	0.33773 (–)	0.68114 (201%)	0.64943 (192%)
Pos. intracloud	0.30005 (–)	2.41890 (806%)	0.65297 (217%)
Pos. cloud-to-ground	0.31975 (–)	3.52504 (1102%)	1.58986 (497%)
Pos. all	0.30650 (–)	2.95904 (965%)	0.85634 (279%)

computed for all the distribution types used in this study and is detailed in *Table 4.4*. A fit quality less than 0.5 is considered acceptable.

Even though the negative distributions have the general shape of a positively skewed unimodal log-normal distribution, the single log-normal distribution is unable to achieve an acceptable quality of fit. The double log-normal distribution approximates the measured distribution much better, with good fits for the distributions where network assigned type classification is assigned.

The Cauchy-gamma distribution is 15 % better than the double log-normal for the negative intracloud distribution and 23 % better for the negative cloud-to-ground distribution.

Since a log-normal distribution is unimodal, the fit quality values above unity for the single log-normal distribution to the bimodal positive cloud-to-ground distribution is understandable. The double log-normal distribution does not describe the positive distributions as well as the negative distributions.

The Cauchy-gamma distribution accurately describes the bimodal distributions, with similar qualities for the positive distributions and negative distributions.

4.4 Fitted results

The Cauchy probability density function, given in *Equation 4.3* is used to describe the cloud-to-ground distribution; and a gamma probability density function, given in

Table 4.5: Parameters of the distributions fitted to the negative probability mass functions (shown in *Figure 4.3*).

	Data set		
Parameter	Intracloud	Cloud-to-ground	All
I_c	0.7990	1.6162	1.5768
I_γ	1.4033	0.7969	0.8276
x_c	13.0030	16.5952	16.5916
k_γ	3.8343	5.7371	5.5724
θ_c	8.0644	10.1083	10.1855
θ_γ	2.5114	2.0073	2.0653

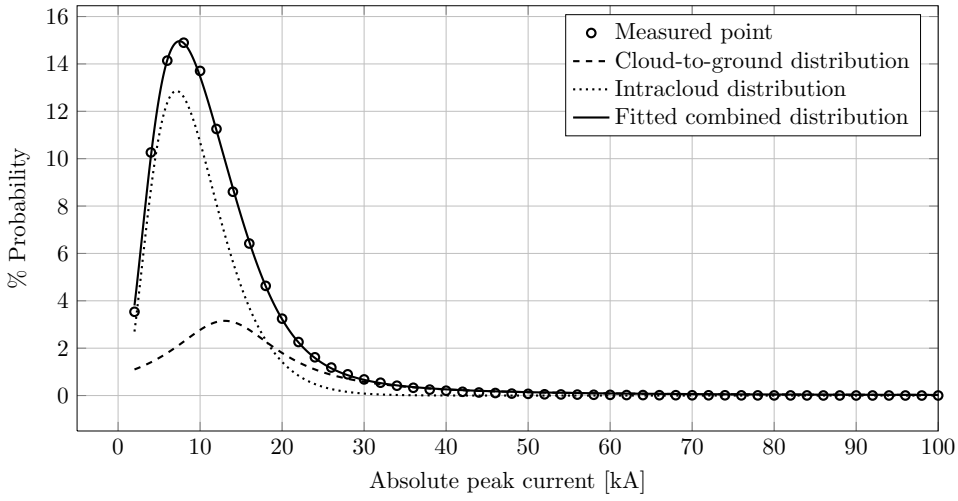
Equation 4.4 is used to describe the intracloud distribution. These two distributions are summed to produce a complex Poisson distribution which is then fitted to the measured distribution using the particle swarm method detailed previously.

As visible in *Figure 4.3* and *Figure 4.4* this method accurately fits the complex distribution to the measured distribution. The parameters of the complex distribution are detailed in *Table 4.5* for the negative distributions and *Table 4.6* for the positive distributions.

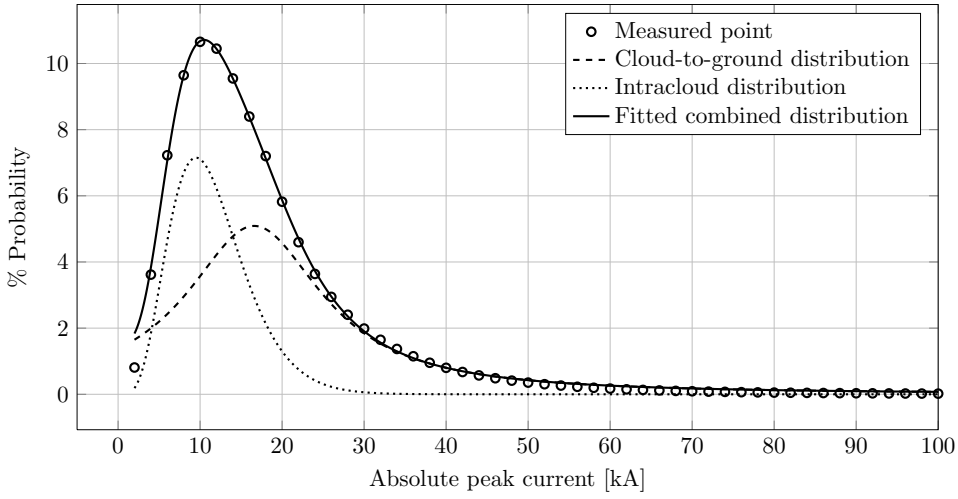
4.4.1 Negative data sets

The intracloud component is the dominant component in the combined distribution fitted to the measured intracloud distribution, shown in *Figure 4.3a*. However, the intracloud component does not completely describe the measured distribution, and there is a small contribution from the cloud-to-ground component, with equal probabilities of occurrence at 18.0 kA. Above 30.0 kA the probability of the intracloud component is almost zero, and the cloud-to-ground component dominates the long tail. The mode of the cloud-to-ground component is at 13.0 kA, while the mode of the intracloud component is at 8.0 kA.

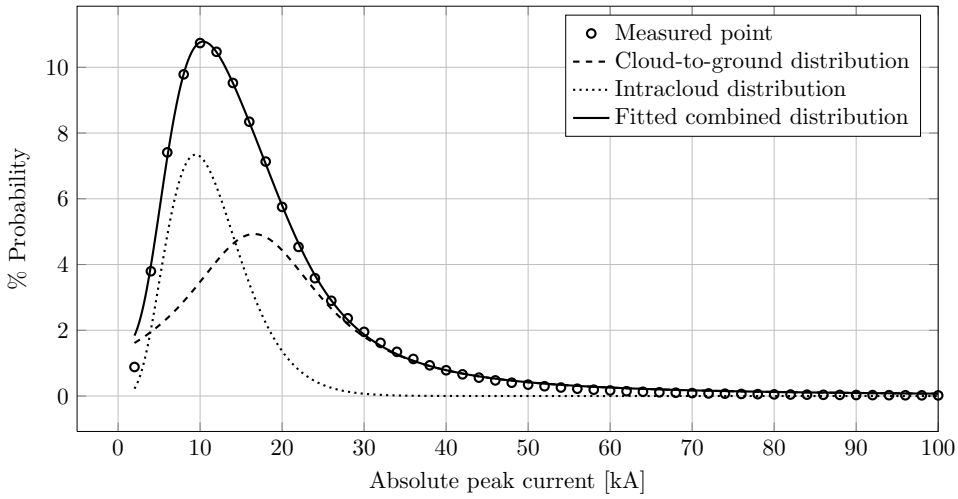
The intracloud component is also present in the cloud-to-ground distribution, shown in *Figure 4.3b*, with a similar shape to the intracloud component in the intracloud distribution, and negligible contribution above 30.0 kA. The cloud-to-ground component is more prominent in this distribution, and has an equal probability of occurrence at 14.0 kA, and unlike in the intracloud distribution the cloud-to-ground component has a significant contribution at the lower currents too. The mode of



(a) Intracloud



(b) Cloud-to-ground



(c) All strokes

Figure 4.3: Probability mass functions fitted to the measured probability mass function for strokes classified as negative.

the cloud-to-ground component in the cloud-to-ground distribution is 3.5 kA greater than the mode of the cloud-to-ground component in the intracloud distribution; the mode of the intracloud component in the cloud-to-ground distribution is only 2.0 kA larger than the intracloud component in the intracloud distribution.

Since the dominant network assigned type classification for negative strokes is cloud-to-ground (97 % of all negative strokes), the distribution where type classification is ignored is representative of the cloud-to-ground distribution, with very little influence from the intracloud distribution.

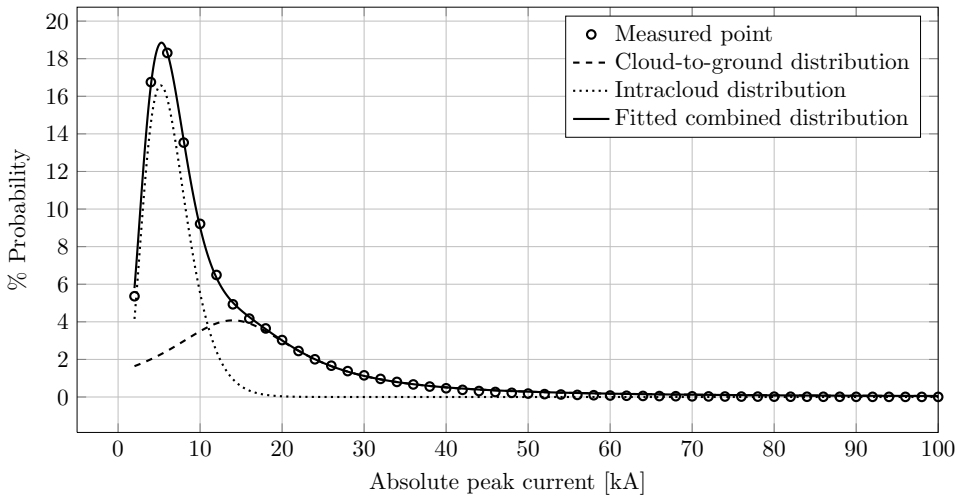
4.4.2 Positive data sets

The positive intracloud distribution shown in *Figure 4.4a* is dominated by the intracloud component, this is similar to the negative intracloud distribution where the intracloud component also dominates. The abrupt change at 15.0 kA is explained by the rapidly decaying probability of the intracloud component with increasing peak current and the increasing probability of the cloud-to-ground component. The intracloud component has a negligible contribution to the combined distribution above 15.0 kA. The mode of the cloud-to-ground component is at 14.0 kA, while the mode of the intracloud component is at 5.0 kA.

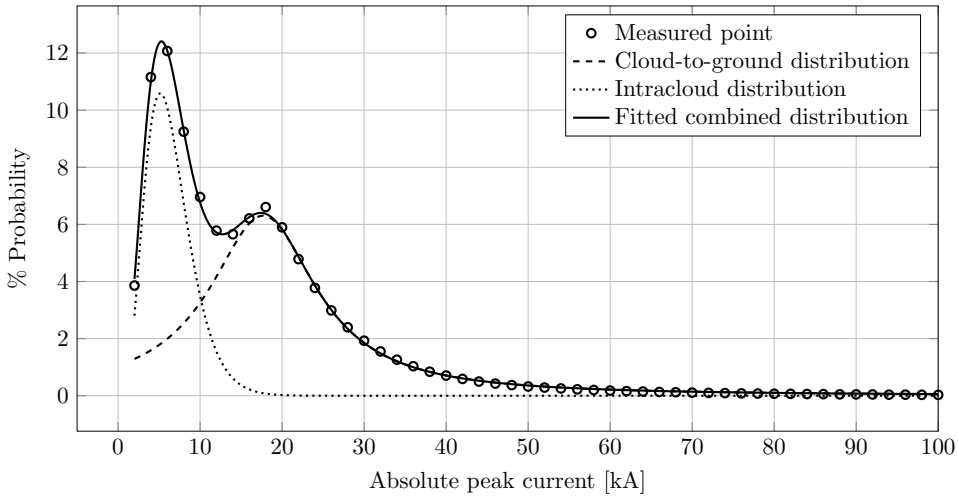
The combined distribution is accurately fitted to the bimodal positive cloud-to-ground distribution, shown in *Figure 4.4b*. The cloud-to-ground component has a similar contribution in the intracloud distribution and the cloud-to-ground distribution where the intracloud component dominates. The second mode is described almost entirely by the cloud-to-ground component, while 83 % of the first mode is

Table 4.6: Parameters of the distributions fitted to the positive probability mass functions (shown in *Figure 4.4*).

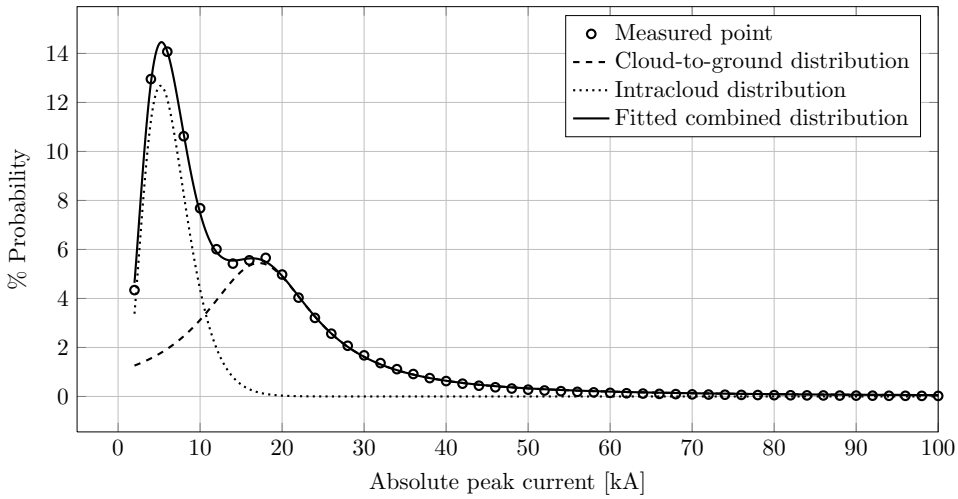
	Data set		
Parameter	Intracloud	Cloud-to-ground	All
I_c	1.2617	1.5730	1.4253
I_γ	1.0926	0.6997	0.8505
x_c	14.0004	17.6756	17.1839
k_γ	5.0713	4.9769	4.9143
θ_c	9.8524	7.9445	8.3340
θ_γ	1.2768	1.2950	1.3239



(a) Intracloud



(b) Cloud-to-ground



(c) All strokes

Figure 4.4: Probability mass functions fitted to the measured probability mass function for strokes classified as positive.

described by the intracloud component. At 5.0 kA, the location of the mode of the intracloud component in the cloud-to-ground distribution is the same as in the intracloud distribution, but the location of the mode for cloud-to-ground component is 3.6 kA larger at 17.6 kA.

Even though there are more positive strokes with a cloud-to-ground network assigned type classification (66 % of all positive strokes), the location of the first mode (intracloud component) remains at 5.0 kA, and the location of the second mode (cloud-to-ground component) is representative of the mode for the cloud-to-ground component in the cloud-to-ground distribution.

Despite the presence of both intracloud and cloud-to-ground components in all of the distributions, the network assigned type classification is not without merit. Each distribution can been seen as the selection for a specific network assigned type classification, as well as the selection against the other network assigned type classification. Hence each distribution created by applying network assigned type classification exhibits a clear bias towards the type selected for when compared against the distribution created by applying the network assigned type classification selected against.

4.5 Comparison to video study

Biagi *et al.* (2007) with the aid of a video camera assessed the classification efficiency of the United States National Lightning Detection Network (NLDN), and *Table 4.7* is his work. The video study was conducted in 2003 and 2004 in Arizona (AZ); and Texas (TX) and Oklahoma (OK). 31 hours of activity are described by Biagi *et al.* (2007), and during this period all the sensors in the region were operational.

For negative strokes, only those with peak currents less than or equal to 10.0 kA were considered, but for positive strokes the ranges between 10.0 kA and 15.0 kA, between 15.0 kA and 20.0 kA, and above 20.0 kA were also examined.

Even though Biagi *et al.* (2007) only considered flashes with single strokes, it is not clear if only strokes with cloud-to-ground network assigned type classification were considered or if network type classification was ignored. There are subtle differences between these distributions as discussed previously in this chapter and *Chapter 3*.

Table 4.7: Classification of small single stroke flash reports from (Biagi *et al.*, 2007, pg 11).

(a) Negative ($|I_p| \leq 10\,\text{kA}$) in AZ and TX-OK

Data set	Night/Day	CG	CB	NL	NLDN
AZ 2003	night	22	4	14	40
AZ 2004	night	8	2	0	10
AZ 2004	day	12	0	18	30
TX-OK 2003	night	14	1	1	6
TX-OK 2004	night	4	1	2	7
TX-OK 2004	day	2	0	9	11
Total	night	48	8	17	63
Total	day	14	0	27	41
Total	–	62	8	44	104

(b) Positive in TX-OK

I_p Categories, kA	CG	CB	NL	Total Count
$0 < I_p \leq 10$	6	80	334	420
$10 < I_p \leq 15$	3	7	71	81
$15 < I_p \leq 20$	2	7	16	25
$I_p > 20$	30	2	13	45
Total	41	96	434	574

CG: Cloud-to-ground channel was observed

CB: Cloud brightened, no channel to ground observed

NL: No change in luminosity was observed

As the fitted distributions have been interpolated to refine the 2.0 kA resolution, two distribution derived measures are compared, in the same ranges, to the video study: the first measure is the maximum of the component ratio on a per point basis and is assigned the symbol $\hat{R}$; and the second measure is a ratio of the areas of each component and is assigned the symbol $\bar{R}$. The maximum of the ratio provides a measure of the upper limit of the ratio between the two distributions, while the area ratio is a measure of the two distributions over the entire range.

A similar ratio is derived from the video observations of Biagi *et al.* (2007): by comparing the number of strokes where a cloud-to-ground channel is observed to either the number of strokes where changes in cloud luminosity were observed (CB:CG) or to all strokes detected by the network that are not cloud-to-ground ((CB+NL):CG). The results of the comparison are shown in *Table 4.8*.

While the video derived ratios are comparable to the component derived ratios, the video study by Biagi *et al.* (2007) for positive strokes consists of 11 hours and 36 minutes spread across four days in a single season, and 19 hours and 32 minutes across two seasons for the negative strokes. The natural ratio between cloud-to-ground and intracloud stroke rates is dependent on a host of meteorological parameters, including storm type, climate cycle and even solar cycle. The decomposed distributions and the corresponding ratios, are for the entire network, sampled for 19 224 hours across 801 days (over four years) which represents a significantly longer sampling period.

Table 4.8: Comparison of stroke count for the intracloud to cloud-to-ground ratio, for different current ranges and component ratios from the fitted distributions.

	Biagi *et al.* (2007)		Decomposed component			
I_p Categories	CB:CG	(CB+NL):CG	$\hat{R}_{\text{cg}}$	$\bar{R}_{\text{cg}}$	$\hat{R}_{\text{all}}$	$\bar{R}_{\text{all}}$
$-10 < I_p < 0$	0.129:1	0.838:1	2.264	1.469	2.403	1.575
$0 < I_p \leq 10$	13.333:1	71.428:1	6.084	3.226	7.442	4.008
$10 < I_p \leq 15$	0.336:1	26.315:1	1.021	0.321	1.332	0.440
$15 < I_p \leq 20$	3.508:1	11.627:1	0.061	0.020	0.090	0.032
$I_p > 20$	0.066:1	0.5:1	0.004	0.000	0.006	0.001

CG: Cloud-to-ground channel was observed
CB: Cloud brightened, no channel to ground observed
NL: No change in luminosity was observed
R_{cg}: Cloud-to-ground distribution
R_{all}: Type independent distribution

Summary

Each distribution may be decomposed into two components: a cloud-to-ground component represented by the Cauchy distribution, and an intracloud component represented by the gamma distribution. The measured distribution is decomposed into the components by fitting the complex distribution to the measured distribution. An efficient numerical method in the form of particle swarm optimisation was used to fit the distributions. Single and double log-normal distributions were also investigated, but the Cauchy-gamma distribution is at least 15 % better.

The location of the mode in the negative distributions only differ by 5.0 kA and the decay of the negative intracloud component is slower than in the positive intracloud components, thus only one mode is visible to a cursory inspection of the negative distributions.

The bimodal positive cloud-to-ground distribution has an intracloud component mode at 5.0 kA and a cloud-to-ground component mode at 17.6 kA, and since these two modes are significantly far apart, they are visible to a cursory inspection of the distribution. The abrupt change in the positive intracloud distribution is explained by the rapid decrease in intracloud component probability and the gradual increase in cloud-to-ground component probability. The numerical method allows for the distribution of the positive cloud-to-ground component to be derived without discarding any of the data based on the proposed re-classification limits.

The next chapter extends the decomposition of the distributions by detailing how the statistical measures vary through the diurnal cycle, and in particular how the mode provides a unique measurement of lightning development.

Chapter 5

Diurnal variations in the statistical parameters of lightning

This chapter begins by introducing the concept of a two parameter measurement, of activity and location, for the various statistical parameters. The temporal development in the diurnal cycle of the activity and location of the mean, median and mode for the various distributions arising from network assigned type classification is given. The contribution of this chapter is in showing the separate temporal development of the intracloud mode and cloud-to-ground mode, despite both originating from the same data set. The diurnal variation is developed into a general charge distribution model for lightning producing clouds in southern Africa.

5.1 The measurement of activity

Figure 5.1 introduces the concept of a two parameter statistical measurement: the location and probability. Thus the mean, median, and mode are regarded as points on the probability distribution with parameters of location (x) and probability (y). If the probability is multiplied by the number of strokes in the data set then the measure is of activity, and since the location parameter is the current at the point then the two parameters are current (x) and activity (y).

For example, the diurnal variation of the mode is now described by the current and activity. If there are many low energy strokes then the mode will have low current but high activity, conversely if there are few high energy strokes the mode will have

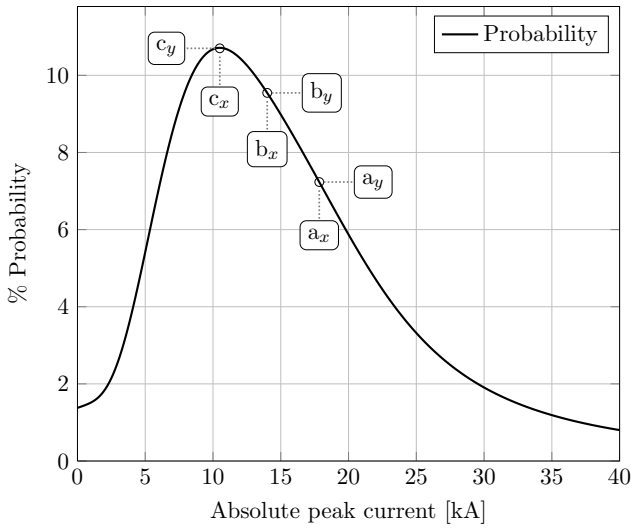


Figure 5.1: Illustration showing the location (x axis) and probability (y axis) of the mean (a), median (b), and mode (c) on a positively skewed probability mass function.

a high current but low activity. These two parameters vary independently of each other and in addition to a measure of activity also provide a measure of energy.

Lightning activity for southern Africa is categorised by the variation of the statistical parameters. The diurnal variation, which is a description of the general case, is developed from data from the South African Lightning Detection Network (SALDN), and the arising data sets are decomposed into their separate components.

5.1.1 Minute of day algorithm

The area in this study, southern African, is two hours east of the universal time meridian, which was previously known as Greenwich Mean Time. International measures of lightning activity, typically obtained from space and Schumann resonance observations, are reported in Co-ordinated Universal Time (UTC), and this convention is continued in this thesis.

Southern Africa is defined as the area south of the northern borders of the Republic of South Africa, including the sovereign nations of the Kingdom of Swaziland and the Kingdom of Lesotho. The lightning detection network is owned and operated by the South African Weather Service (SAWS).

To convert to South African Standard Time (SAST) two hours must be added to the UTC time base. South Africa does not have a daylight saving scheme therefore

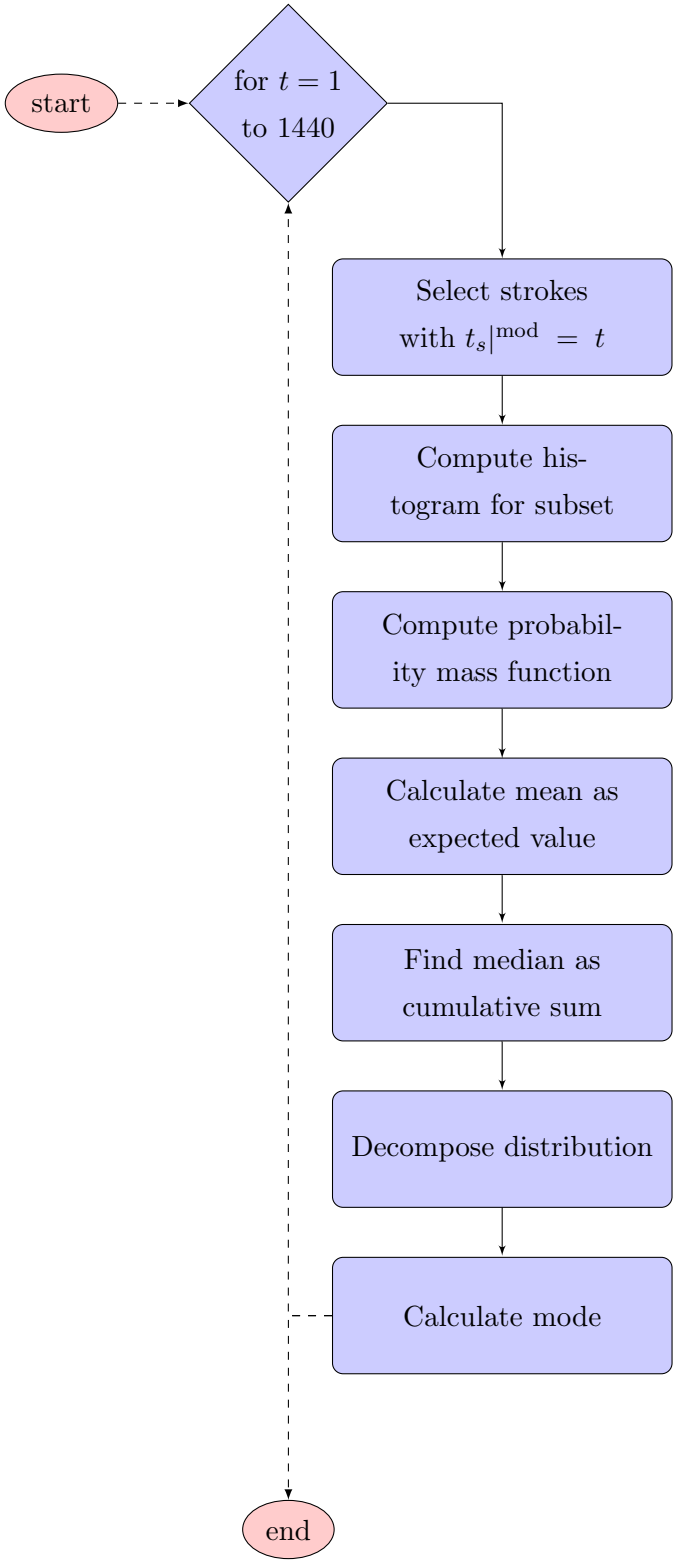


Figure 5.2: Algorithm for the calculation of the temporal variation of the mean, median, and decomposed modes. The resolution is a minute, and the loop iterates through all the minutes in a day/diurnal cycle.

no seasonal adjustments should be made to the time base.

The diurnal cycle is the variation of parameters through a 24 hour period, and starts at midnight (00:00 UTC) until midnight the following day (24:00 UTC). Diurnal cycles typically recur every calendar day. In order to calculate the variation of statistical parameters through the diurnal cycle, a histogram for each minute of the diurnal cycle is computed. From this histogram the statistical parameters are computed.

Figure 5.2 details the specific implementation of the minute of the day algorithm, where there are 1440 min in a day. For each minute only the strokes occurring in that minute are selected and from this subset a histogram is computed. The histogram is then converted into a probability mass function, as described in *Chapter 3*, from which the mean and median are computed. The probability mass function is then decomposed into the separate intracloud and cloud-to-ground components and the modes of the components are found.

5.1.2 Inspection and decomposition

The probability mass functions for the negative data sets are unimodal, but it was shown previously that, regardless of the network assigned type classification, there is an intracloud component and a separate cloud-to-ground component. In the negative data sets, the modes are too close to each other to be measured directly, and so each histogram or probability mass function must be decomposed in order to measure the parameters of the components.

This is not the case for the positive data sets, where the location of modes are sufficiently far apart that the location of each mode may be measured through inspection of the distribution. However, measurements of the activity of each mode, made directly on the distribution, inherently contain components of the other component and are thus influenced by the diurnal variation of both components. Measurements made on the decomposed components are independent of the other component in both current and activity.

Figure 5.3 shows the diurnal variation of the second mode of the positive cloud-to-ground data set measured with the inspection method and the decomposition method. The series described by black circles in *Figure 5.3* is the mode measured by inspection and the grey crosses in the same figure is the mode measured through decomposition.

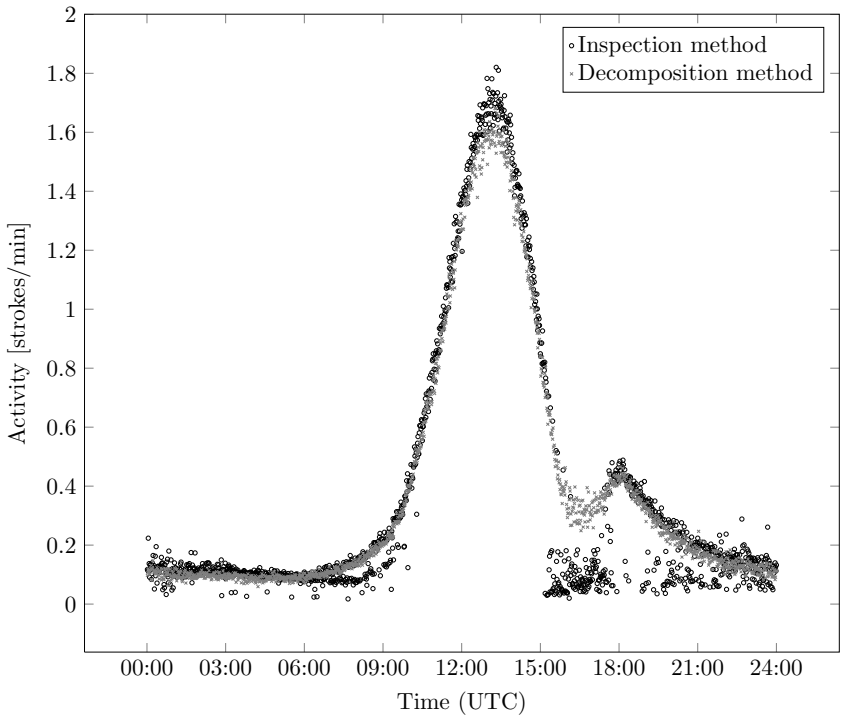


Figure 5.3: The second mode of the positive cloud-to-ground data set measured by inspection and through decomposition.

The temporal variation in the mode as measured by the two different methods is not significantly different, but in the period of reduced activity between 15:00 UTC and 18:00 UTC and subsequently in the final decay of activity towards the end of the diurnal cycle the inspection method fails to detect the mode. The source code from the implementation of both the inspection and decomposition algorithms are included in *Appendix B*

The mean and median measurements are made with the inspection method so that stroke rates may be compared with other international studies, although the diurnal variation of these parameters has not been documented. Except for *Figure 5.9*, the modal measurements are made on decomposed distributions.

5.1.3 Window

The minute of the day algorithm only selects strokes that occurred within a specific minute as data for the histogram. The histogram could be interpolated since a minute interval is a very fine resolution, however this process artificially filters the probability mass function as well as increases the apparent number of samples in that set. This is not a consistent mathematical operation, and hence is not used.

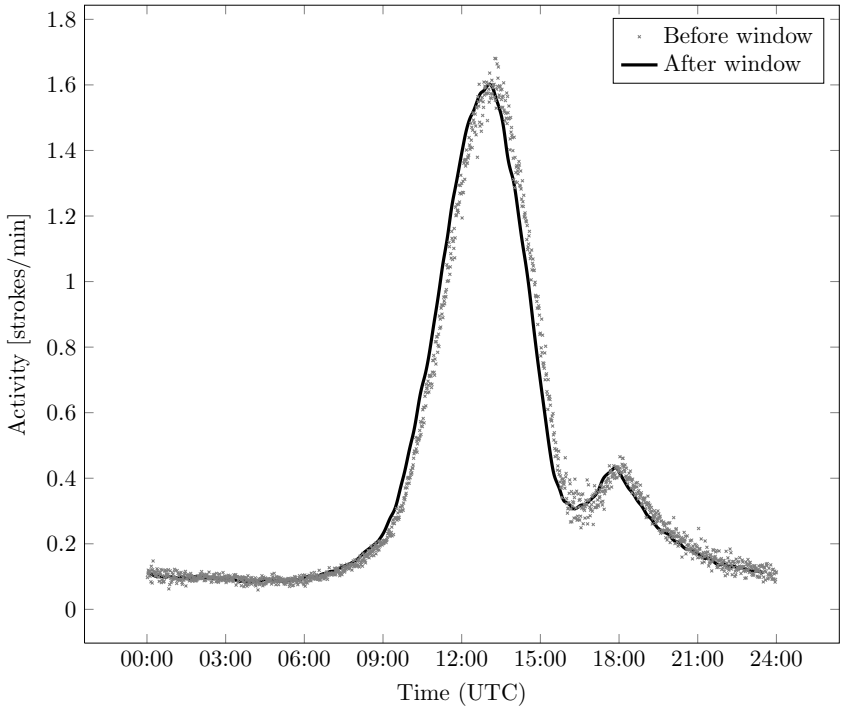


Figure 5.4: The second mode of the positive cloud-to-ground data set showing the effect of the windowing operation.

The temporal developments are minimally smoothed through the convolution of a window with the time series. The convolution of a window of equal values with a time series is a computationally efficient form of computing the moving average of the time series. A 10 minute wide Hann function is used as the window instead of a set of equal values, so that the central point in the window dominates. The definition of a Hann function is:

$$w(n) = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi n}{N_w - 1}\right)\right) \tag{5.1}$$

where

$w(n)$ = Window function

N_w = Number of points in window

The Hann function has a unity peak in the centre of the window and a sinusoidal decay to the edges of the window, which are zero. The effect of the convolution of the window with the time series is similar to a moving average except that the central point has the greatest contribution and the outer points have no contribution. Thus greater sensitivity is achieved over a simple moving average. The advantage of using the Hann function over other window functions is that very little aliasing is introduced by this function (Oppenheim *et al.* (1998)).

The area of the window is normalised so that the data are not scaled—this is equivalent to dividing the sum of a set of points by the number points in calculating a moving average. The window area is normalised by the following equation:

$$w_a(n) = \frac{w(n)}{\int_{-\infty}^{\infty} w(u)du} \tag{5.2}$$

where

$w_a(n)$ = Area normalised window of $w(n)$

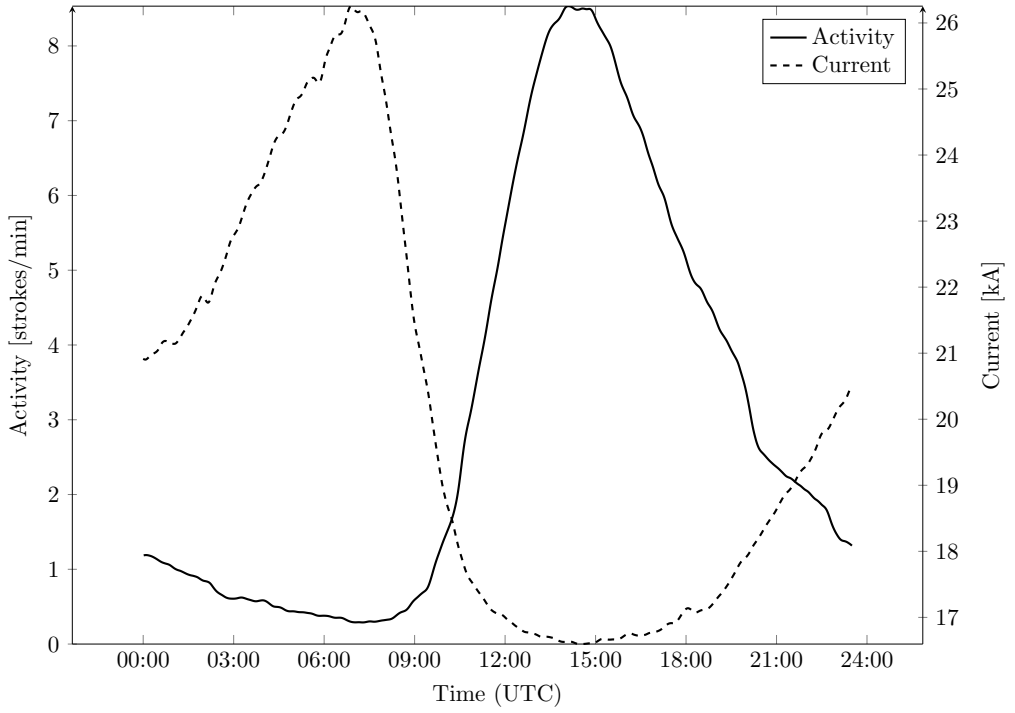
The performance of the window is shown in *Figure 5.4*, where the points show the data before convolution with the window, and the solid curve is the data after convolution with the area normalised window. The convolution result follows the trend of the data, although there is a small phase shift.

5.2 Mean of the cloud-to-ground data set

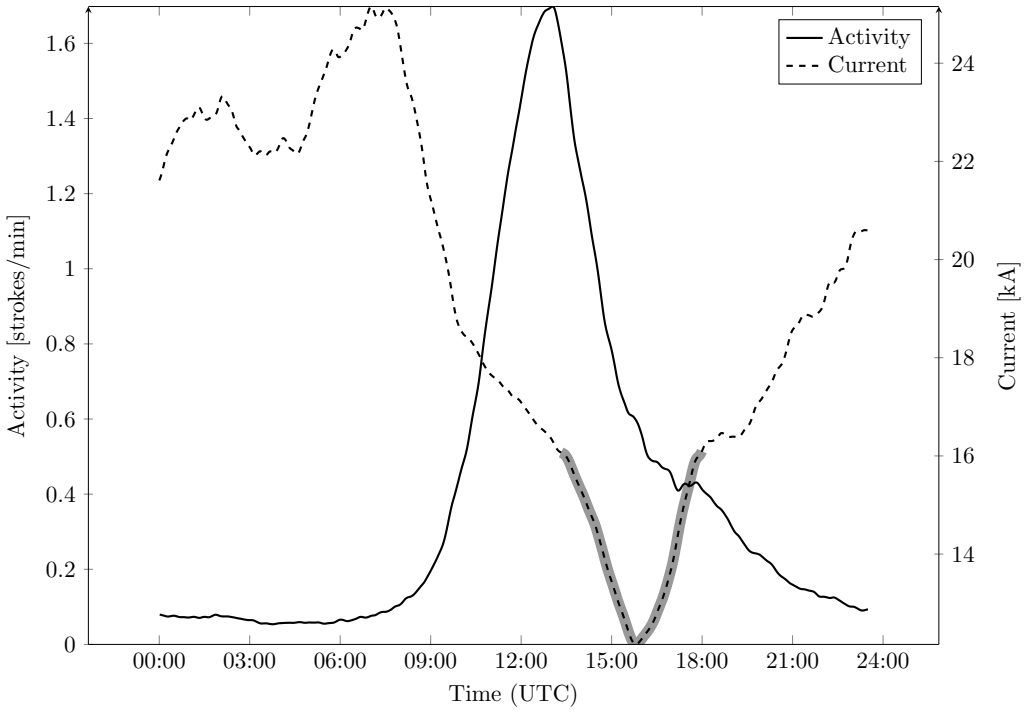
The diurnal cycle is visible in the activity of the negative cloud-to-ground data set, displayed in *Figure 5.5a*, and correlates well with African diurnal cycles measured from Schumann resonances by Price *et al.* (2004). Similar diurnal cycles are also visible in North American data, as reported by Zajac & Rutledge (2001). The sun rises between 03:00 UTC and 06:00 UTC and the surface of the earth begins to warm. There is a delay between the increase of lightning activity and sun rise due to the heat capacity of the surface of the earth; the mass of earth must heat up before thermal energy can be transferred into the air at the surface. Lightning activity persists after sun set as latent energy from the previous day is transferred from the surface of the earth into the atmosphere. From the lightning minimum at 09:00 lightning activity rapidly increases and is at the maximum at 14:00 UTC, and from 15:00 UTC the gradual decay in lightning activity begins with the rate of decay slower than the rate of increase.

The positive cloud-to-ground data set has not had the Cummins *et al.* (1998) reclassification limit of 10.0 kA or the Cummins & Murphy (2009) reclassification limit of 15.0 kA applied.

The diurnal cycle is also visible in the activity of positive cloud-to-ground data set, displayed in *Figure 5.5b*. Unlike the activity of the negative cloud-to-ground mean there is no decay from the positive cloud-to-ground activity in the preceding day. Even though fewer strokes with positive cloud-to-ground classifications are observed,



(a) Negative Strokes



(b) Positive Strokes

Figure 5.5: Temporal development of activity and current at the mean, from the set of strokes classified as cloud-to-ground. Note the confined minimum in the positive cloud-to-ground current (highlighted).

the increase in activity of the positive mean begins an hour before the negative mean due to the substantial presence of intracloud strokes in this data set.

Continuing the trend the positive cloud-to-ground mean activity peaks an hour earlier, but then decays quicker than the negative cloud-to-ground mean. There is an abrupt change in the rate of decay between 15:00 UTC and 18:00 UTC, which corresponds to the setting of the sun and a reduction in the incident energy to the storm cell.

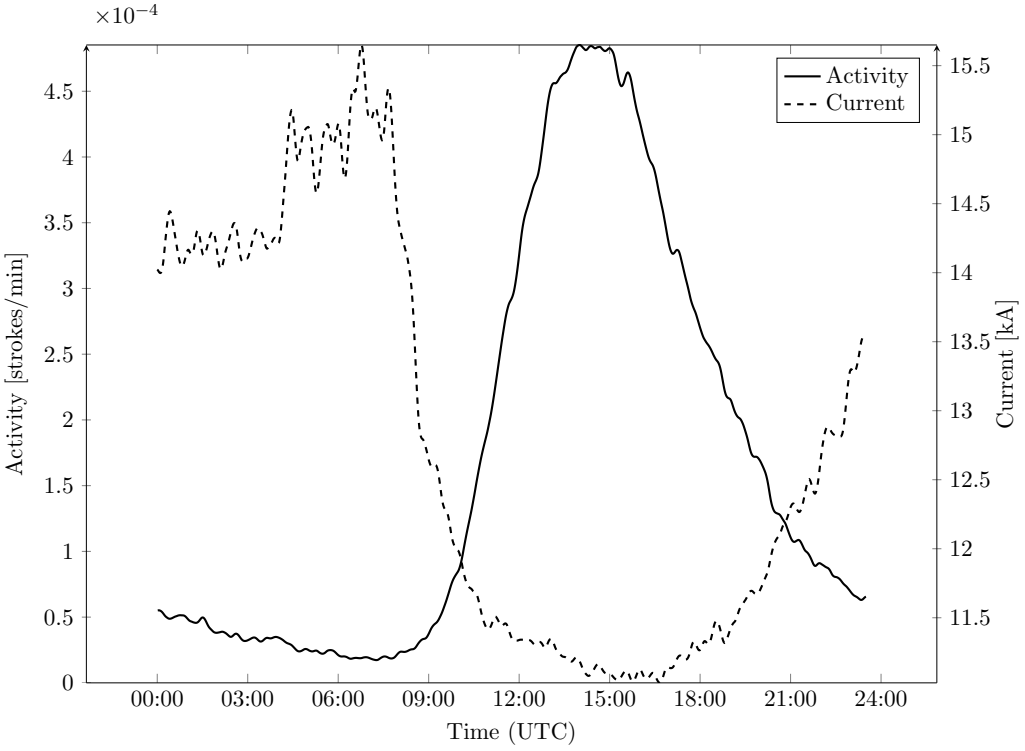
Although the current at the mean of the positive cloud-to-ground data set begins the diurnal cycle with a larger magnitude than the current at the mean of the negative cloud-to-ground data set, the two current temporal profiles peak at 07:00 UTC and decay at comparable rates. The rate of decay of the positive peak current current profile changes as the positive activity peaks, resulting in a temporally confined minimum, as highlighted in *Figure 5.5b*. The minimum arises because the positive cloud-to-ground process has almost completely halted and the misclassified intracloud strokes dominate. Hence the mean does not provide a consistent measure for this data set without applying inflexible reclassification limits.

5.3 Mean of the intracloud data set

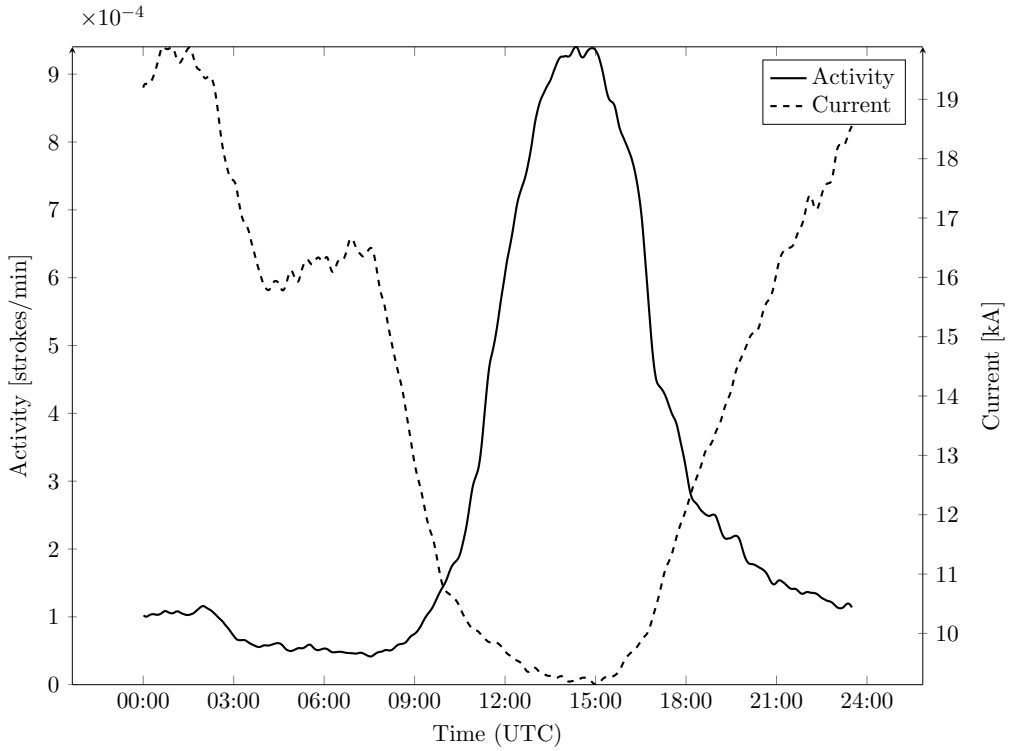
The diurnal cycle for negative intracloud strokes is shown in *Figure 5.6a*, and the shape of the activity cycle closely resembles that of the negative cloud-to-ground diurnal cycle. The measure of activity at the negative intracloud mean is in the order of 10^{-4} strokes/min because this is not a measure of all the activity, but of the number of strokes observed at the mean of the distribution for the various minutes of the day.

Unlike the negative cloud-to-ground current profile, the negative intracloud current profile at the mean starts at 90 % of the peak and then gradually rises to the peak at 09:00 UTC; and as the activity increases the current decreases.

The temporal development of the positive intracloud activity, shown in *Figure 5.6b* is similar to the negative intracloud activity: lightning activity begins to increase after 09:00 UTC and peaks at 13:30 UTC. The activity of the positive intracloud mean begins to decay at 15:00 UTC and decays quicker than the activity at the negative intracloud mean.



(a) Negative Strokes



(b) Positive Strokes

Figure 5.6: Temporal development of activity and current at the mean, from the set of strokes classified as intracloud.

The current at the positive intracloud mean starts at 95 % of the peak, and peaks at 01:00 UTC, at 02:00 UTC there is a rapid reduction which corresponds to a small decrease in activity. The current remains relatively constant until the activity increases.

5.4 Median of the cloud-to-ground data set

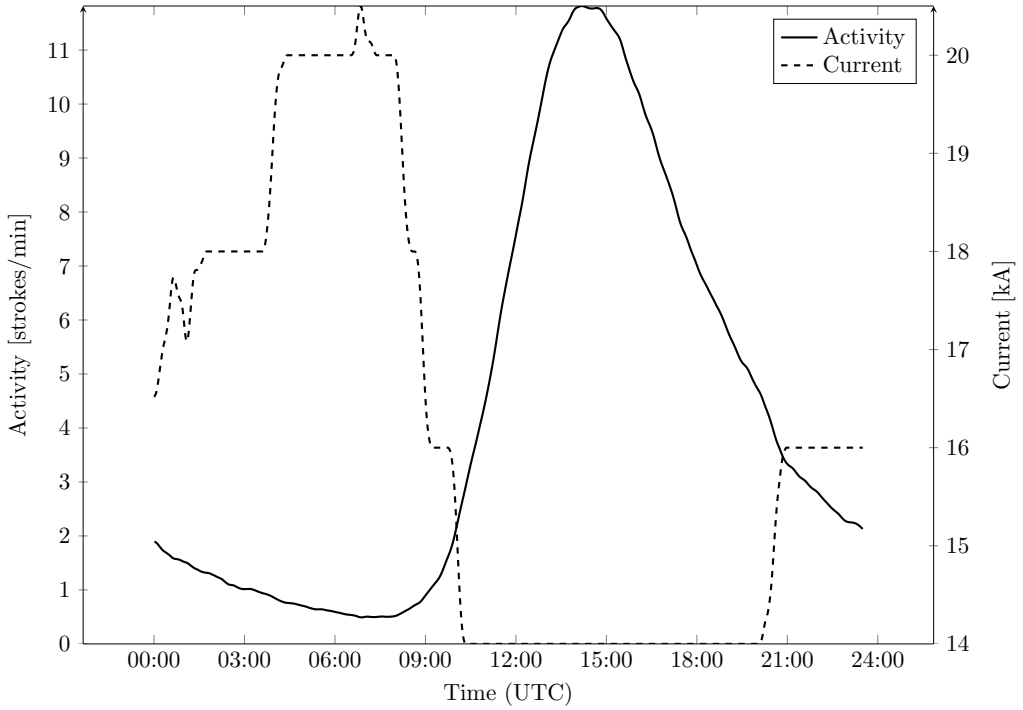
The temporal development of the negative cloud-to-ground median through the diurnal cycle is shown in *Figure 5.7a*. This temporal development of the median is the same as the mean, where the latent energy from the previous day decays to a minimum in lightning activity between 06:00 UTC and 09:00 UTC. After 09:00 UTC lightning activity rapidly increases to a peak at 14:00 UTC. The decay in lightning activity from the peak to the end of the day is gradual and at the same rate as the mean.

The algorithm for calculating the location of the median iterates through points on the cumulative probability mass function until the cumulative probability is greater or equal to 0.5. Since the resolution of the peak current is limited to 2.0 kA and the histogram is not interpolated, as explained previously in *Section 5.1.3*, there is an inherent resolution limitation in the calculation of the median. This is visible in the temporal development of the current at the median in *Figure 5.7a* as a set of discrete steps in the curve. Despite the resolution limitations of this measurement the shape of the curve is similar to the current at the mean in *Figure 5.5a*.

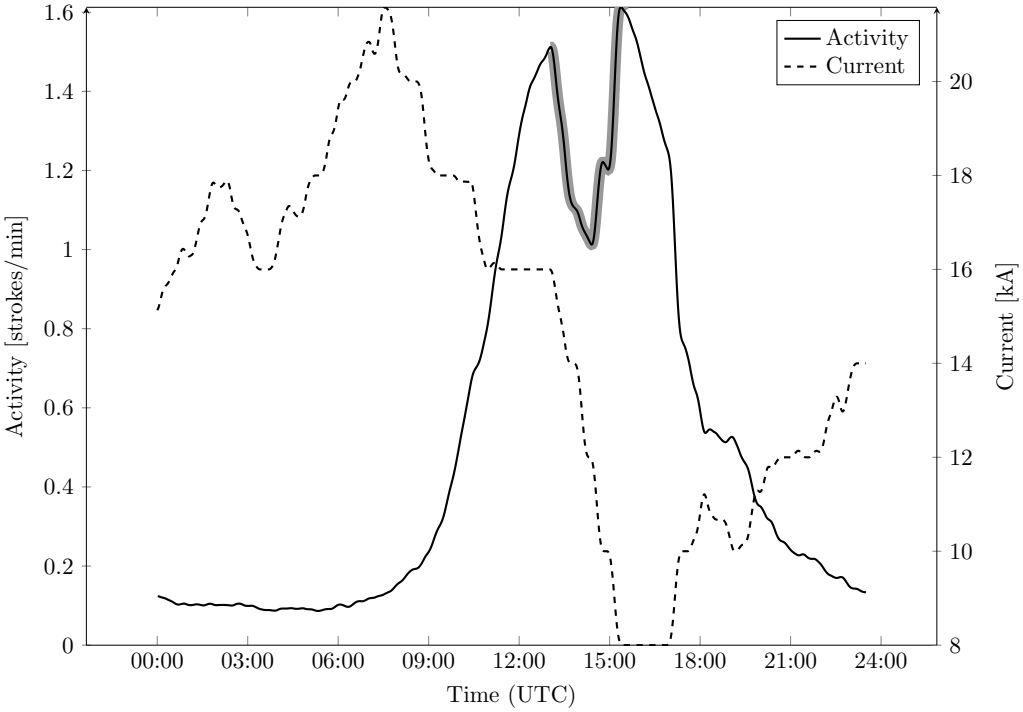
The temporal development of the median for the positive cloud distribution is shown in *Figure 5.7b*. The highlighted section in *Figure 5.7b* shows the trend of the peak of the activity at the mean inverting, and is at a local minimum when the negative cloud-to-ground median is at a maximum.

The inversion in the trend occurs because the relative number of misclassified intracloud strokes increases with respect to the number of correctly classified cloud-to-ground strokes, the median moves into the local minimum between the two modes of the bimodal positive cloud-to-ground probability density function (see *Figure 3.3* and *Figure 4.4b*). This is also supported by the reduction of the current at the median, indicating that the median is moving towards the origin.

The first peak in the double peaked development occurs at the same time as the peak of the positive cloud-to-ground mean, shown in *Figure 5.5b*, and the second



(a) Negative Strokes



(b) Positive Strokes

Figure 5.7: Temporal development of activity and current observed at the median, from the set of strokes classified as cloud-to-ground. Note the inversion in the peak of the positive median, as highlighted.

peak occurs at the same time as the minimum of the positive current at the mean profile, shown in *Figure 5.5b*. Hence the current at the positive mean and activity at the positive median is influenced by the presence of the misclassified intracloud component and neither the mean nor median provides a consistent measure of the temporal development.

5.5 Median of the intracloud data set

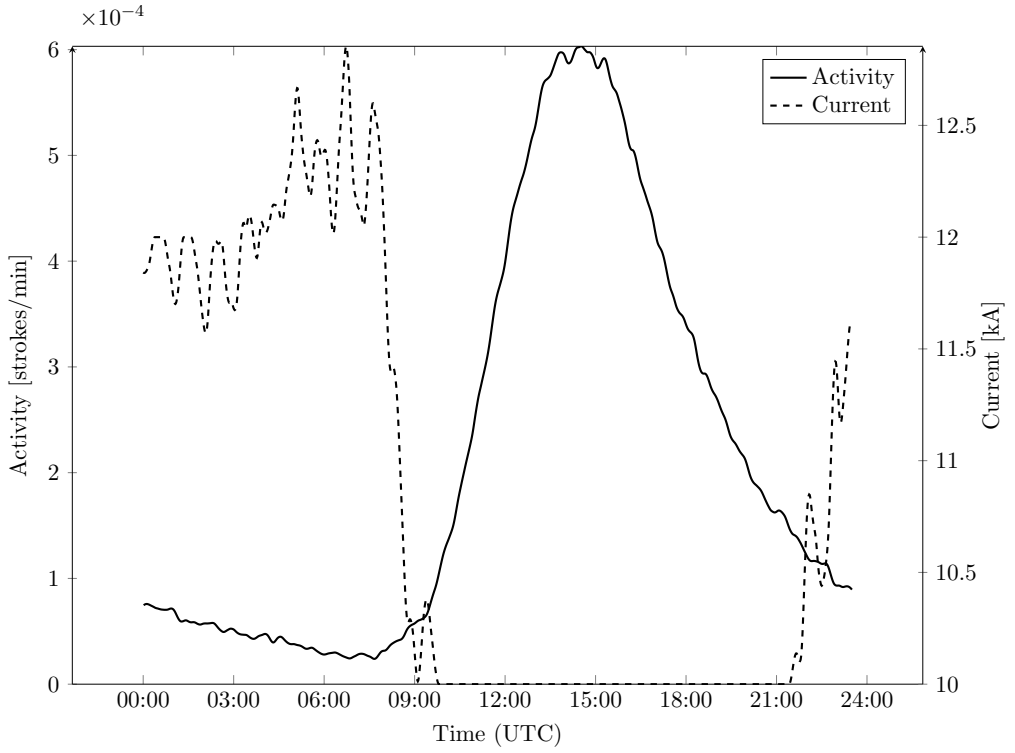
The temporal development of the negative intracloud activity is shown in *Figure 5.8a*. The intracloud diurnal cycles follows the same pattern as the activity of the intracloud and cloud-to-ground data sets for the negative mean as well as the cloud-to-ground data set for the negative median.

The activity minimum occurs between 06:00 UTC and 09:00 UTC, from then on there is a rapid increase to the peak at 14:00 UTC. The activity of the median of negative intracloud activity begins to decay at 15:00 UTC, and the rate of decay slows at 21:00 UTC.

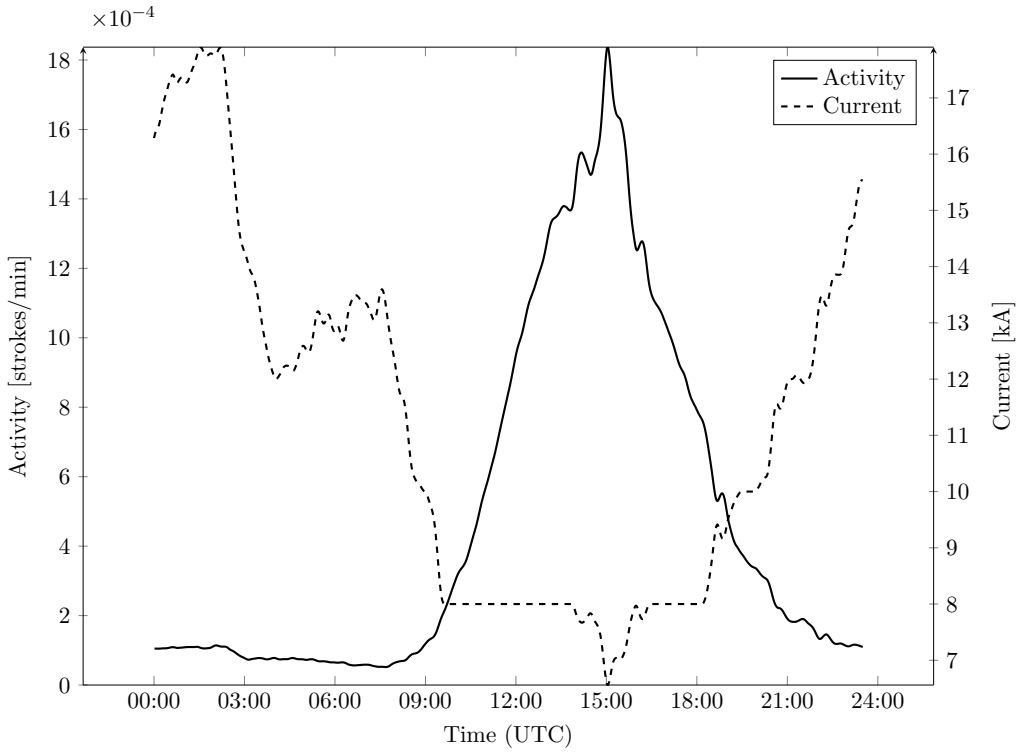
The diurnal cycle for the median of the positive intracloud data set is shown in *Figure 5.8b*. This cycle has a similar trend to the mean of the positive intracloud data set and the peak at 15:00 UTC in *Figure 5.8b* corresponds to the second peak of the activity at the median of the positive cloud-to-ground data set shown in *Figure 5.7b*.

The peak in activity at the median is not as broad as the peak in activity at the mean, and arises because the median is on the limit between adjacent histogram bins. For example, a small change in activity is artificially enhanced because the median is now measured at the 7.0 kA bin and not the 8.0 kA bin. For this reason, median based measurements are not a true reflection of lightning activity and are inherently flawed.

The current at the median for the positive intracloud data set is also similar to the current at the mean for the positive intracloud data set: the cycle starts at 88 % of the peak which occurs at 02:00 UTC and then rapidly decays to an intermediate level of 66 % of the peak and then in line with the behaviour observed in the previous data sets the current reduces as activity increases and increases as activity decreases.



(a) Negative Strokes



(b) Positive Strokes

Figure 5.8: Temporal development of activity and current observed at the median, from the set of strokes classified as intracloud.

5.6 Mode of the cloud-to-ground data set

The mode shown in *Figure 5.9* is obtained though inspection of the distribution, while the modes in *Figure 5.10a* and *Figure 5.10b* are from the decomposed components of intracloud and cloud-to-ground distributions. The decomposition algorithm and results are discussed in detail in *Chapter 4*.

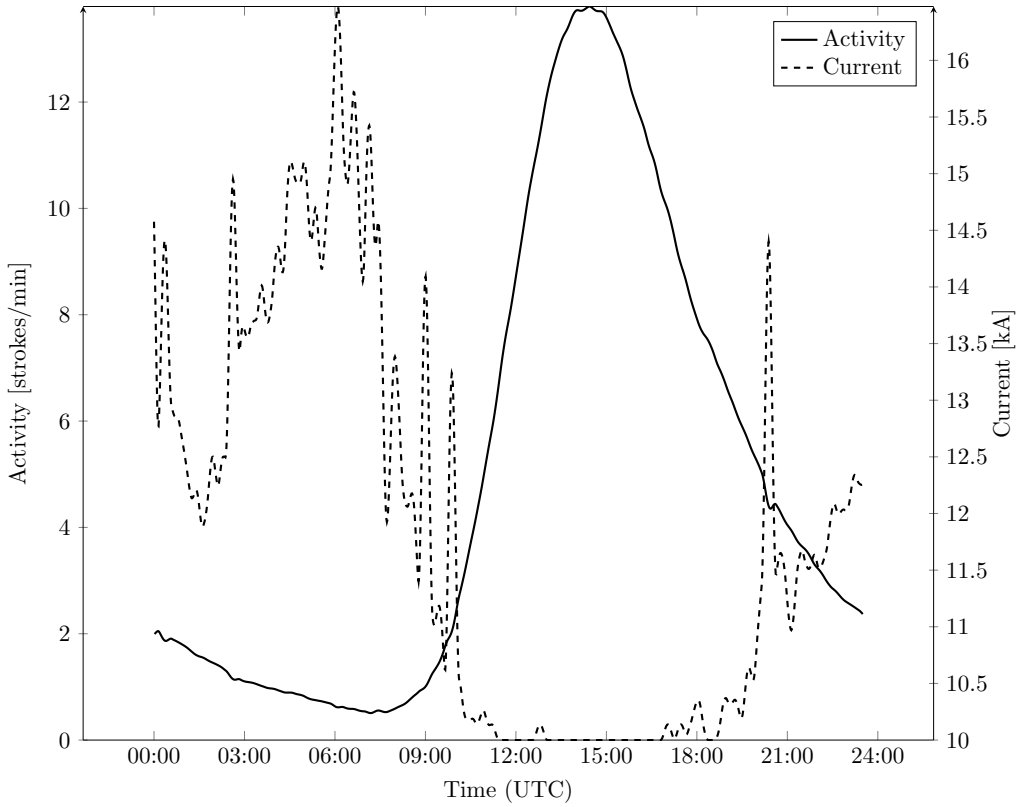
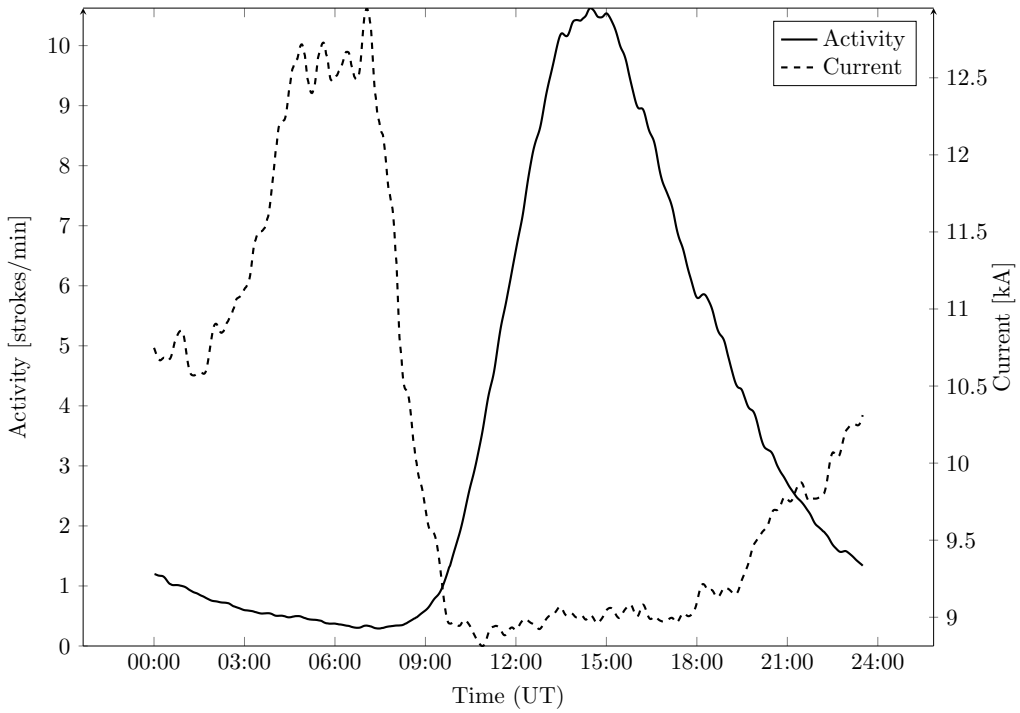
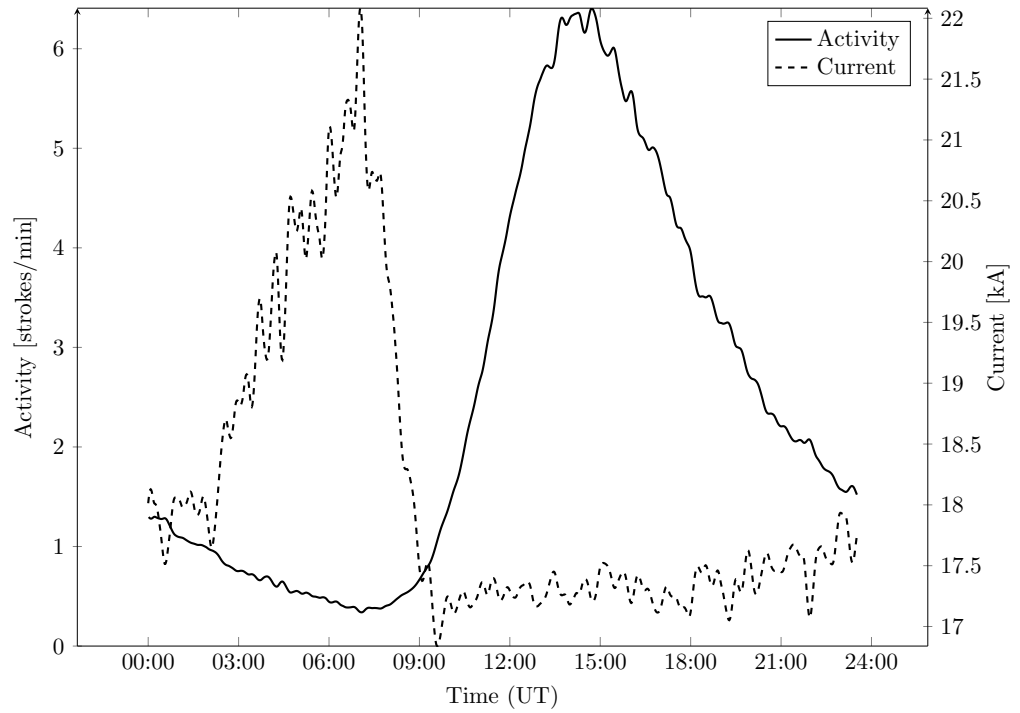


Figure 5.9: Temporal development of activity and current, observed through inspection, at the mode of the negative cloud-to-ground data set.

The second mode in the negative cloud-to-ground data set, measured through inspection of the data set, occurs between 20.0 kA and 65.0 kA, and the level of activity at the second mode is an order of magnitude smaller than the first mode. Any variation in the second mode is a result of the signal to noise ratio changing. As activity increases the number of strokes per histogram increases and so the profiles of histograms in high activity periods are better defined. Random permutations in high activity periods occur at higher currents where stroke counts are less and the histogram is poorly represented, thus peaks that the algorithm is sensitive to are in fact random noise. There is no value in this measurement and it is not presented.



(a) Mode of intracloud component



(b) Mode of cloud-to-ground component

Figure 5.10: Temporal development of activity and current at the decomposed modes, from the set of strokes classified as negative cloud-to-ground.

The intracloud component in the negative cloud-to-ground data set is shown in *Figure 5.10a* and is computed from the parameters of the gamma distribution used to model the distribution of the intracloud process. As a result of this the mode is precisely determined from the parameters of the fitted distribution and is not confined to the 2.0 kA resolution of the mode measured through inspection. The same process is used for the Cauchy distribution used to model the cloud-to-ground process.

The diurnal variation of the intracloud component is similar to the diurnal cycle of the inspected mode for the negative cloud-to-ground data set, and starts at 10 % of the peak. The quantisation of the current is not visible due to the method used to compute the current at the mode, and the peak activity is 75 % less.

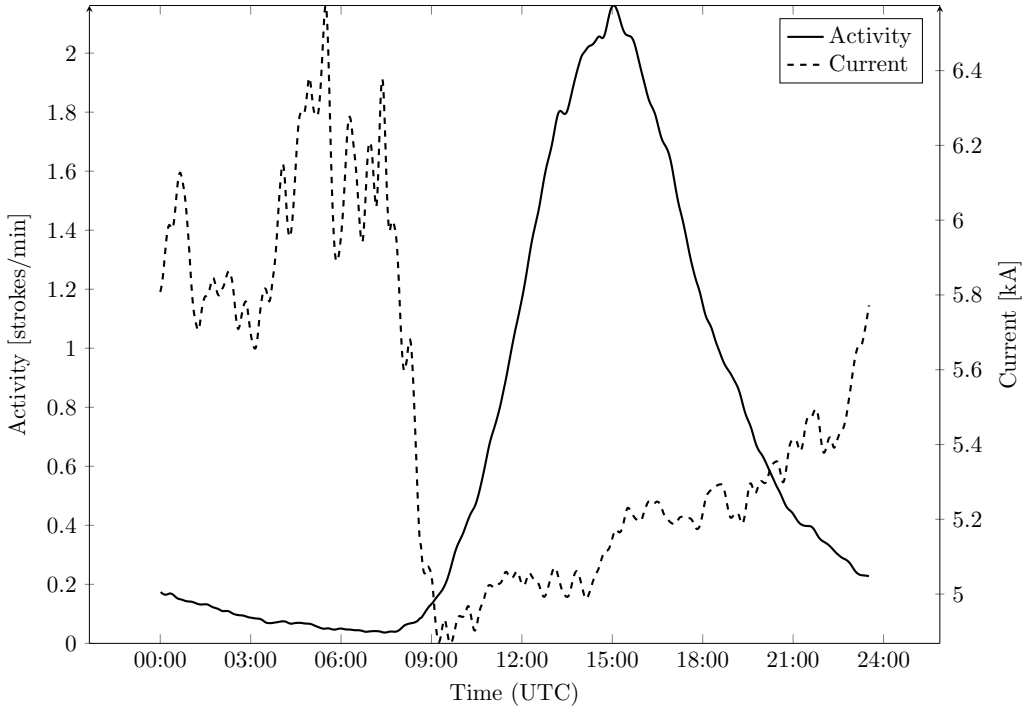
The mode of the cloud-to-ground component also has a similar pattern, however the activity at the beginning and end of the diurnal cycle is larger than the intracloud component at 20 % of the peak. The decay in activity from the peak to the end of the diurnal cycle is slower than the intracloud process.

Unlike the median and mean parameters, the first mode of the positive cloud-to-ground data set exhibits the diurnal cycle present in the mean and median of the negative data sets. The first mode of the positive diurnal cycle is similar to the negative cloud-to-ground diurnal cycles, but starts at only 10 % of the peak where as the negative cycles start at 20 % of the peak.

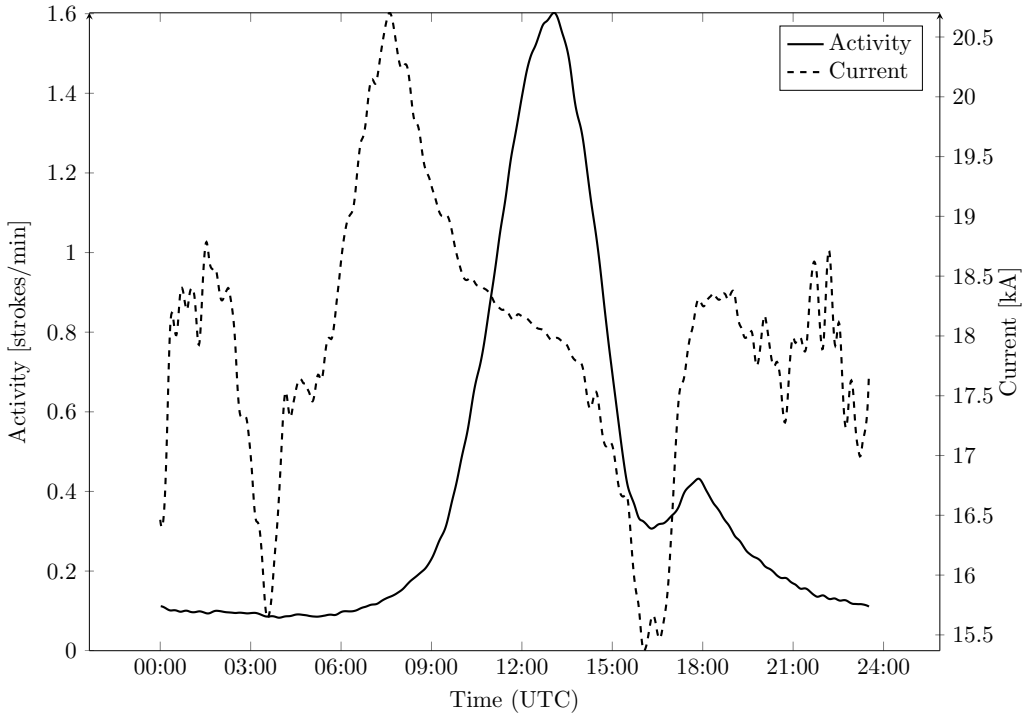
Minimum activity occurs between 06:00 UTC and 09:00 UTC and then increases at the same rate as the negative cycles to a peak at 14:00 UTC. The decay from the peak occurs at the same rate as the negative cycles until 18:00 UTC, and then the first positive cloud-to-ground mode decreases quicker, and finishes the diurnal cycle at 10 % of the peak whereas the negative cycles finish at 20 % of the peak.

The current at the first positive cloud-to-ground mode, shown in *Figure 5.11a*, is less than the Cummins *et al.* (1998) reclassification limit of 10.0 kA and the Cummins & Murphy (2009) reclassification limit of 15.0 kA. As discussed in *Chapter 4*, this section of the data set is mostly comprised of misclassified intracloud strokes. Intracloud activity closely follows negative cloud-to-ground activity and so correlation between these two data sets is expected.

The second mode of the positive cloud-to-ground data set, shown in *Figure 5.11b*, is not in the noise band because the activity levels are similar to the activity levels of the first mode, and the current is always above the Cummins *et al.* (1998) limit



(a) Mode of intracloud component



(b) Mode of cloud-to-ground component

Figure 5.11: Temporal development of activity and current observed at the mode, from the set of strokes classified as positive cloud-to-ground.

of 10.0 kA. The current at the second mode is only below the Cummins & Murphy (2009) limit of 15.0 kA between 02:00 UTC and 06:00 UTC, during which the activity level does not change noticeably.

5.7 Mode of the intracloud data set

The contribution of the cloud-to-ground component calculated through the decomposition of the intracloud data set, presented previously in *Chapter 4*, is limited and the dominant component in this data set is the intracloud component. *Figure 5.12a* shows the diurnal variation of the mode of the intracloud component and the diurnal cycle is similar to the negative diurnal cycles presented earlier.

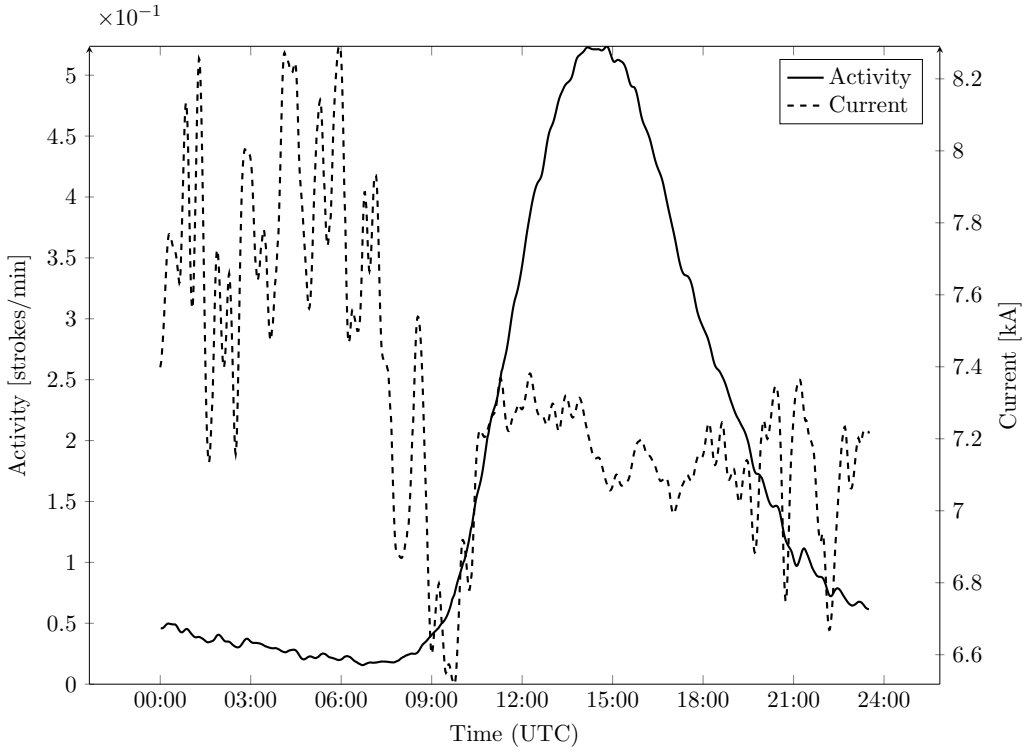
The diurnal variation of the cloud-to-ground mode, as shown in *Figure 5.12b*, is not pronounced, since the contribution of the cloud-to-ground component in this data set is limited. The stroke rates for this component are an order of magnitude smaller than the intracloud component.

This is not the case for the positive intracloud data set where the cloud-to-ground component provides a significant contribution to the entire data set.

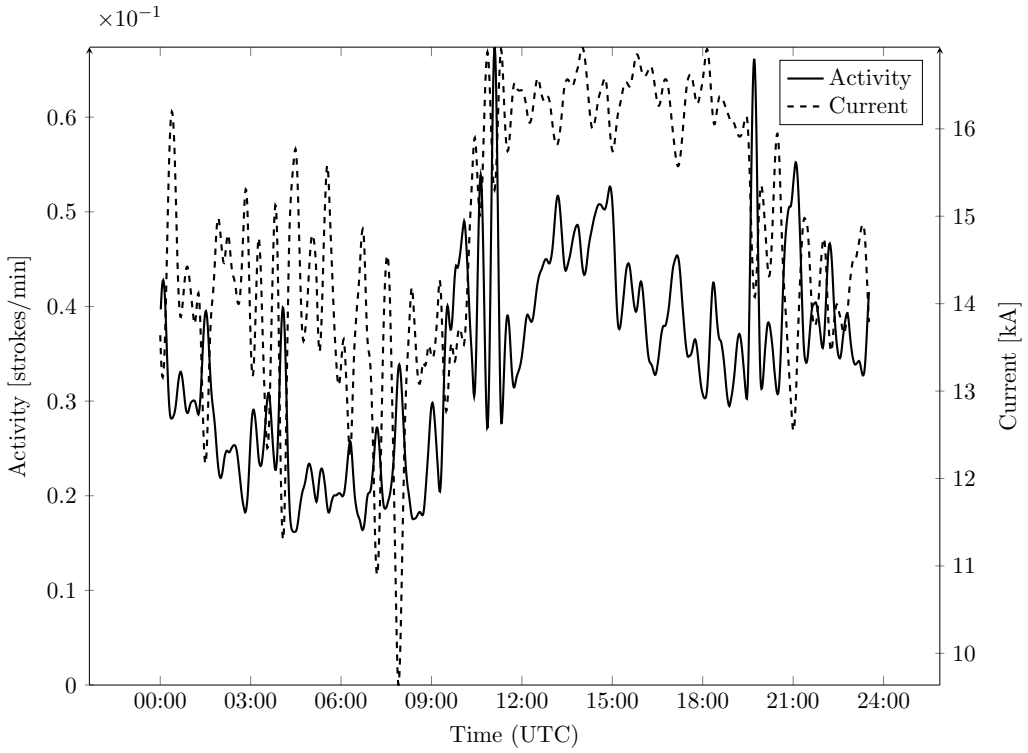
Figure 5.13a follows the diurnal trend established previously and correlates well with the intracloud components of the cloud-to-ground data sets. The increase to peak begins after the period of depressed lightning activity at 09:00 UTC and then peaks at 15:00 UTC. The rate of decay is the same as the other intracloud components.

The positive cloud-to-ground component, shown in *Figure 5.13b*, is similar to *Figure 5.11b* because it too is double peaked, and the peaks occur at similar times, with the reduction in activity between the two peaks corresponding to the maximum in intracloud activity.

The current profile at the cloud-to-ground mode of the positive intracloud data set has a minimum at 09:00 UTC and then steadily increases to a maximum at the end of the day. This maximum propagates into the start of the diurnal cycle.

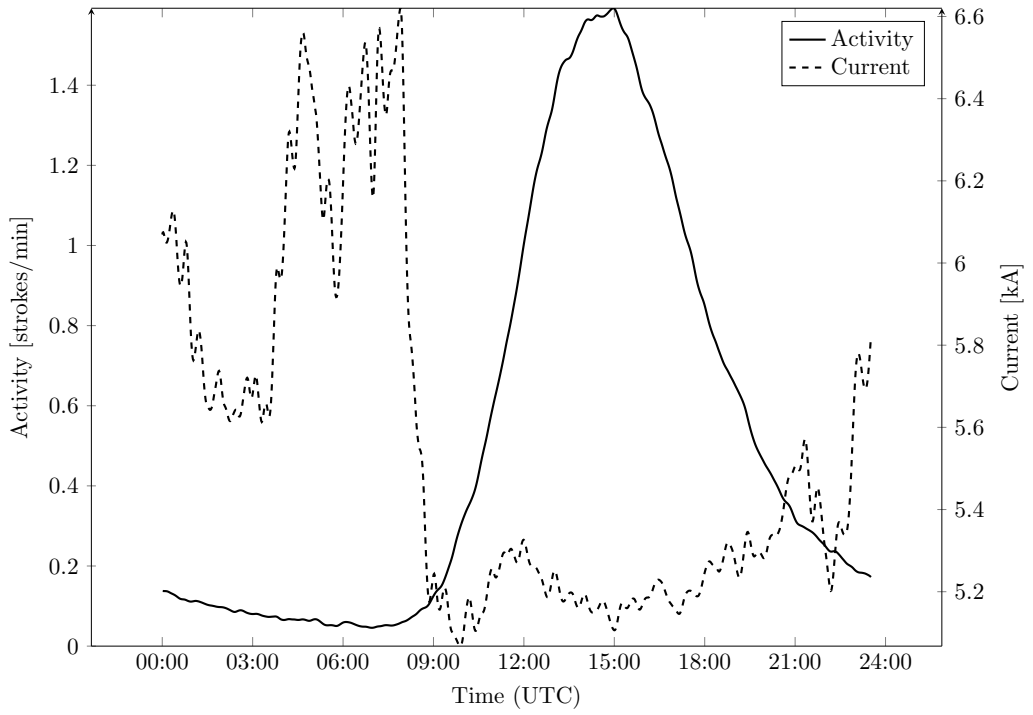


(a) Mode of intracloud component

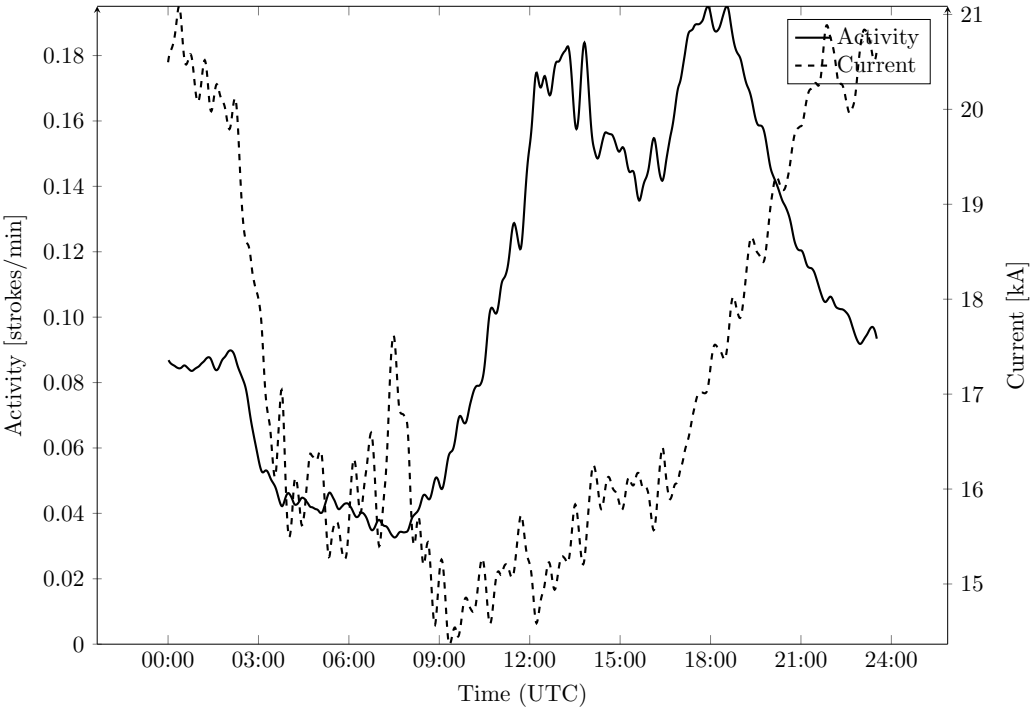


(b) Mode of cloud-to-ground component

Figure 5.12: Temporal development of activity and current observed at the mode from the set of strokes classified as negative intracloud.



(a) Mode of intracloud component



(b) Mode of cloud-to-ground component

Figure 5.13: Temporal development of activity and current observed at the mode from the set of strokes classified as positive intracloud.

5.8 Relationship between activity and current

There is a strong inverse relationship between negative cloud-to-ground activity at the mean and peak current magnitude at the mean, and is also observed in the intracloud data sets as well as the decomposed components for the modes. The inverse trend between activity and current is explained by the fact more activity implies less available charge per stroke and hence lower peak current.

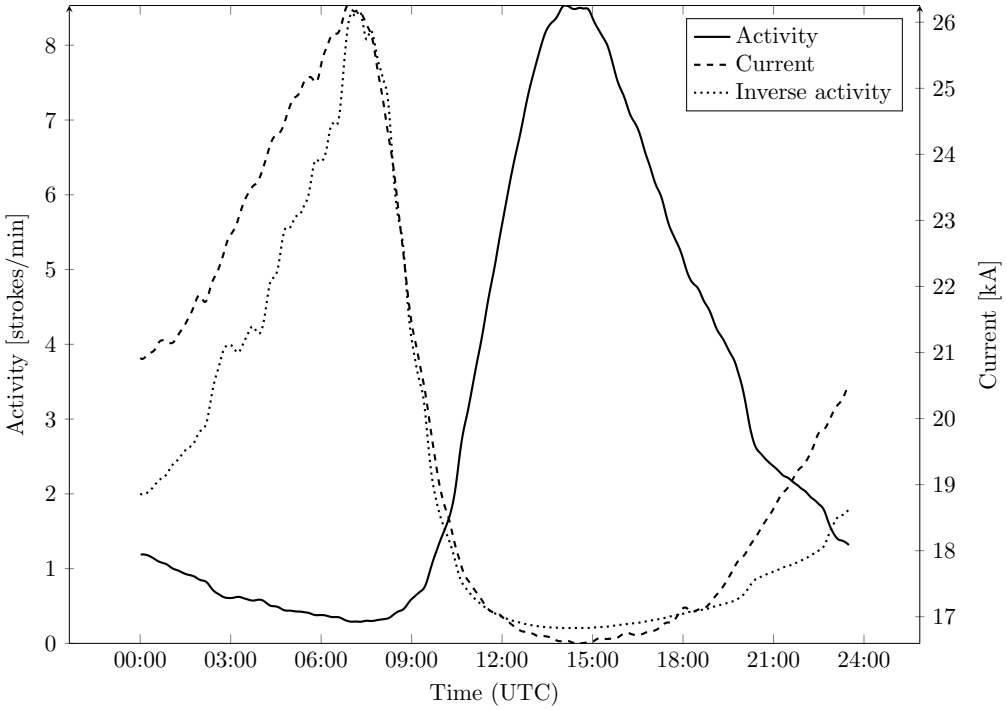


Figure 5.14: Temporal variation of activity, scaled inverse activity, and current at the mean of the negative cloud-to-ground data set.

The inverse relationship between the negative cloud-to-ground activity and peak current magnitude at the mean, shown in *Figure 5.14*, is approximated by the following empirical relationship, with parameters $p = 2.8$ and $o = 16.5$:

$$i(t) = \frac{p}{a(t)} + o \tag{5.3}$$

where

$i(t)$ = Peak current magnitude at diurnal time t [kA]
$a(t)$ = Activity at diurnal time t [strokes/min]
p = Peak for inverse relationship [kA^2]
o = Offset for inverse relationship [kA]

The inverse activity-current temporal relationship is also present in the median analysis, within the 2.0 kA resolution of the current profile.

The inverse relationship is present in the decomposed modal analysis, particularly within the decomposed intracloud and negative cloud-to-ground components. The parameters of the modal relationships are similar to the mean relationship and values for p are between 1.7 and 2 and for o are between 9 and 16. This adds additional confidence to the measure of temporal variation of the decomposed components.

5.9 Charge distribution model

In the initial phase of the development of lightning producing clouds the vertical extent of the cloud is limited and the charge distribution through the cloud follows the tripole model. This is shown in the first panel of *Figure 5.15*. As the rate of convection increases and the vertical extent of the cloud expands, additional charge layers form. There are additional layers of charge in regions where there are strong up drafts(Rakov & Uman (2003)).

The second panel in *Figure 5.15* illustrates the complex charge structure in a mature lightning producing cloud; these complex charge structures have been measured with balloon based electric field measurements (see Cooray (2003) and Rakov & Uman (2003)).

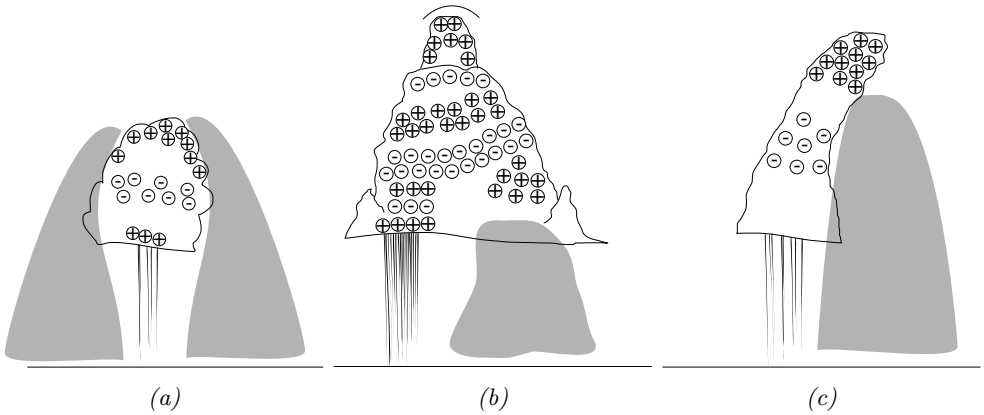


Figure 5.15: Charge distribution within various stages of lightning producing clouds: (a) active, (b) mature, (c) terminal. The shaded volumes indicate likely paths of cloud-to-ground leaders from positive charge centres.

The lower charge layers present an attachment opportunity to leaders originating from upper layers of charge and thus intracloud lightning activity increases at the expense of cloud-to-ground lightning activity.

As storm activity progresses the lower layers of charge are neutralised by intracloud lightning, negative cloud-to-ground lightning and charge on precipitation. Eventually all that remains of the active lightning producing cloud is the vertical distribution of condensed water. The decaying storm structures collapse since the convection process which drove the vertical development has stopped. These final stage clouds have the charge structure that is illustrated in the third panel of *Figure 5.15*.

There are few, sparsely distributed, negative charge centres and positive charge confined to the top of the cloud. Leaders originating from the upper positive charge centres are able to propagate towards ground without attaching to cloud borne charge.

Summary

This chapter presented a two parameter measure of the statistical parameters—current and activity—at the parameter under analysis. An algorithm that computes the statistical parameters for each minute of the day was presented, and the diurnal variation of the statistical parameters was detailed.

It was shown in this chapter that due to the presence of misclassified components in all of the data sets, the mean and median measures and the associated diurnal variations are inherently flawed. The application of fixed reclassification limits is not valid because the natural lightning phenomenon is not bounded by similar limits. Thus while the reclassification limit may be valid for a vast majority of the misclassified strokes in the positive cloud-to-ground data set, the operation is invalid on a significant section of the correctly classified component.

Though the modes may be measured through inspection the decomposition of the data sets for each minute of the day, into intracloud and cloud-to-ground components, provides a separable and robust measure of the components and their associated diurnal variations.

The diurnal variation of the real positive cloud-to-ground component provides an insight into the general charge structure of lightning producing clouds in south-

ern Africa. The development of the vertical extent of a lightning producing cloud effectively shields the upper positive charge layer from the ground.

When negative stroke rates are at their highest—when convection is the most active and the vertical extent of the cloud is the highest—the positive cloud-to-ground lightning activity is significantly depressed.

The next chapter presents the results of modal analysis on select storm days, and extends the conclusions of this chapter to smaller time scales.

Chapter 6

Temporal development of the statistical mode for select days

In this chapter the mode is measured for select storm days and the diurnal variation presented in the general case is extended to specific days. The contribution of this chapter is in showing the applicability of the modal measurement on a limited data set such as a single day.

6.1 Selection and analysis of storm days

The annual lightning cycle in southern Africa peaks between the late spring and summer months, which is between November and April in the southern hemisphere. Two days from the same storm season were selected to account for short term variability, and then two other days from subsequent storm seasons were selected from a data set of 801 days. There is no limitation on application of the method, and it is possible to run the analysis on every single day, however it is not possible to present the entire data set here. The general diurnal variation, which uses all the data, is presented previously in *Chapter 5*.

The data is only analysed in terms of the variation of the mode since unlike the mode the mean and median measures are compromised by the misclassification of stroke type, as discussed previously in *Chapter 4* and *Chapter 5*.

Each day is described by three measures:

- the second mode of the negative cloud-to-ground data set, which represents negative cloud-to-ground strokes;

- the first mode of the positive cloud-to-ground data set, which represents intra-cloud strokes; and
- the second mode of the positive cloud-to-ground data set, which represents positive cloud-to-ground strokes.

The day of 4 March 2010 has the highest stroke rate recorded by the South African Lightning Detection Network (SALDN) since inception of the network. The high lightning activity is due to the presence of a sustained low pressure system over central southern Africa.

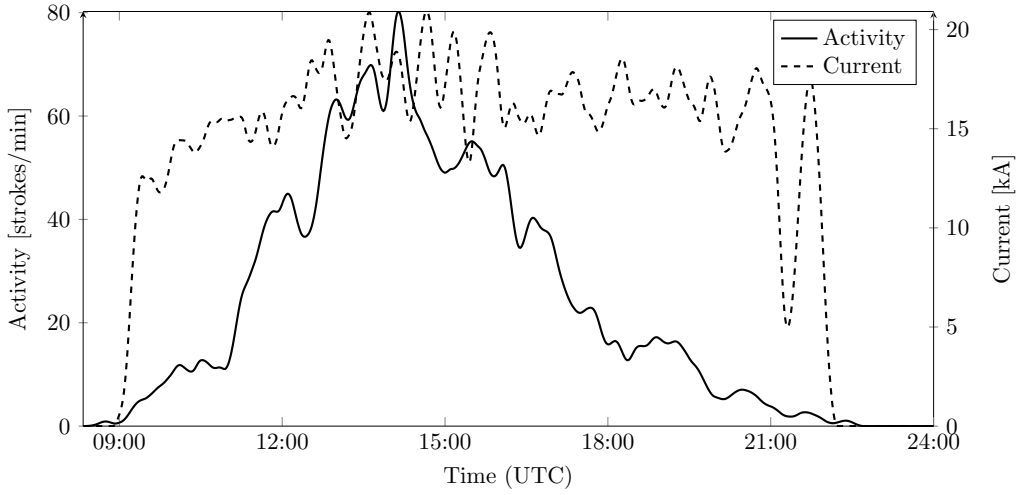
As with the diurnal variation presented in *Chapter 5*, time in this chapter is given in Co-ordinated Universal Time (UTC) to aid international comparison. Two hours are added to UTC to get the South African Standard Time (SAST) as the Republic of South Africa is two hours east of the meridian. There is no day light saving scheme and no seasonal adjustments are made to the time base. For consistency with the diurnal variations presented previously, the same window 10 minute wide Hann window is used.

The analysis occurs over the entire detection region of southern Africa, no attempt is made to mask out the enclave of Kingdom of Lesotho which is located to the east of central South Africa or Swaziland which is on the north eastern border. The detection efficiency for the enclave of the Kingdom of Lesotho and Swaziland is the same as for surrounding regions (Gill (2008)).

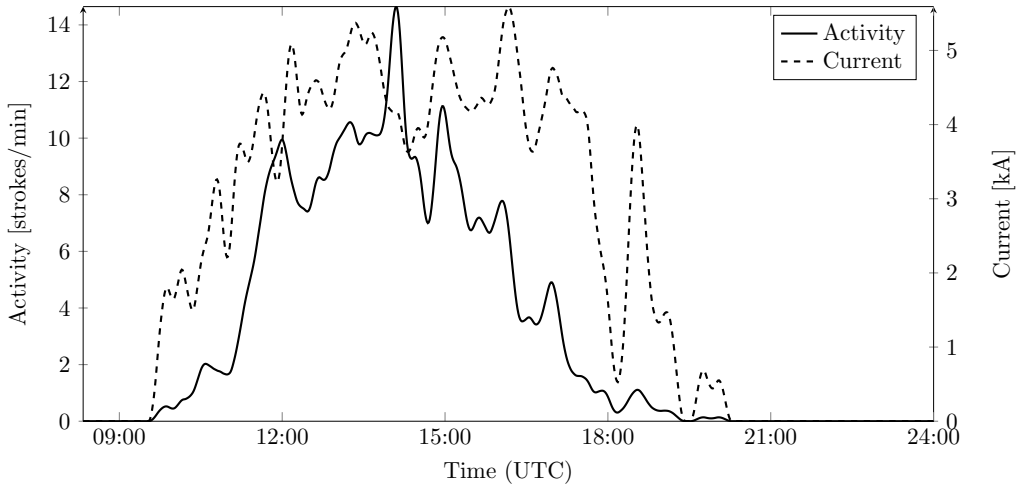
6.2 17 November 2007

The lightning activity on 17 November 2007 is shown in *Figure 6.1*, consisting of the negative cloud-to-ground mode, and the positive intracloud and cloud-to-ground modes. The oscillations in the current at the modes at the beginning and end of the set is due to zeros introduced by no measured activity and the response of the Hann window.

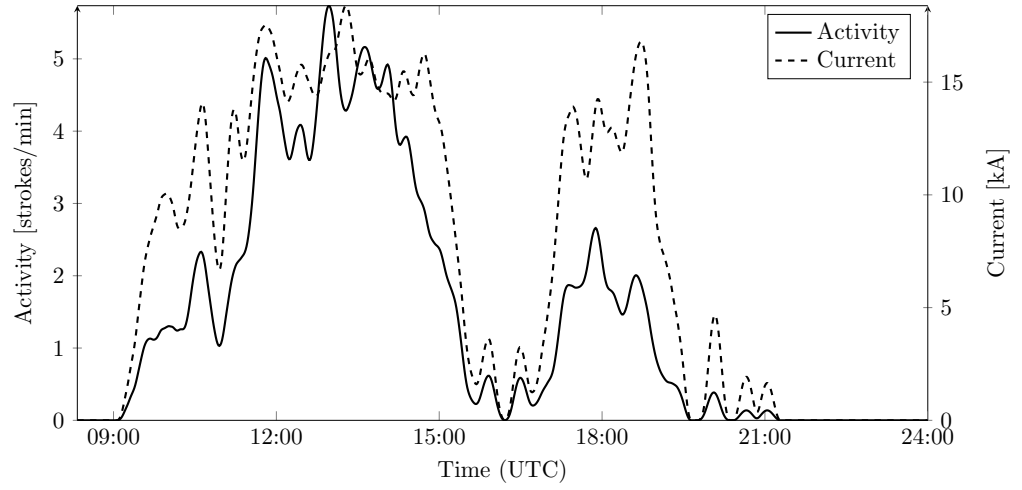
The time scale in *Figure 6.1* starts at 09:00 UTC since there is no lightning activity before 09:00 UTC on the day. After 09:00 UTC there is an gradual development of lightning activity until 11:00 UTC, after which there is a rapid increase in lightning activity to the peak at 14:00 UTC.



(a) Negative cloud-to-ground strokes



(b) Positive intracloud strokes



(c) Positive cloud-to-ground strokes

Figure 6.1: Temporal development of activity and current observed at the modes, for the storms on 17 November 2007.

The development of storms on 17 November 2007 continues at 15:00 UTC which is marked by the reduction in positive cloud-to-ground activity as intracloud activity dominates. The convective processes stop around 17:00 UTC when lightning originating from the upper layer of positive charge begins to attach to ground. The second period of positive lightning is characterised by little intracloud lightning and negative cloud-to-ground lightning.

The development of storms on 17 November 2007 is illustrated schematically in *Figure 6.2*:

***Figure 6.2* - A:** Initially there is moisture in the atmosphere, but little energy to drive convection and so the vertical extent of the clouds is limited.

***Figure 6.2* - B:** As the convection process deepens, the clouds grow vertically. The charge separation mechanisms accelerate and more lightning is produced. The additional layers of charge that are added screen the upper positive charge layer from the ground.

***Figure 6.2* - C:** Finally, as the convection process halts, the upper levels of the cloud are no longer supported and the clouds begin to dissipate and collapse. The lower charge layers are depleted and the upper positive charge layer is now a source of cloud-to-ground lightning.

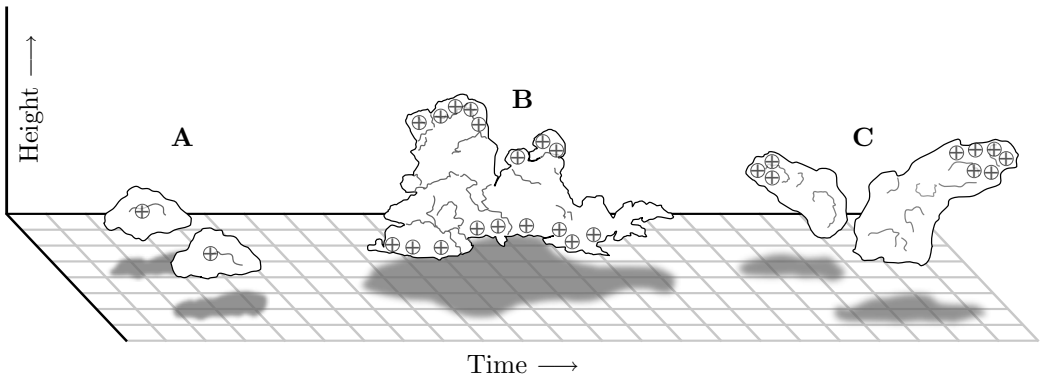


Figure 6.2: Illustration depicting the development of clouds and likely locations of positive charge on 17 November 2007.

6.3 26 November 2007

The day of 26 November 2007 begins with lightning activity already established. Lightning activity before sunrise is driven by the presence of latent energy, absorbed by the surface of the earth in preceding days, being transferred into the atmosphere through convection and thus the charge separation mechanisms that source the charge for the lightning phenomenon keep functioning.

The development of the positive intracloud lightning component, shown in *Figure 6.3*, begins at sunrise. As the sun rises, incident energy is absorbed by the pre-existing cloud layer and not the surface of the earth, resulting in more energy on the surface of the clouds which is not transferred into the cloud. Deep convective processes that could form strong up drafts, resulting in multiple charge layers, are depressed.

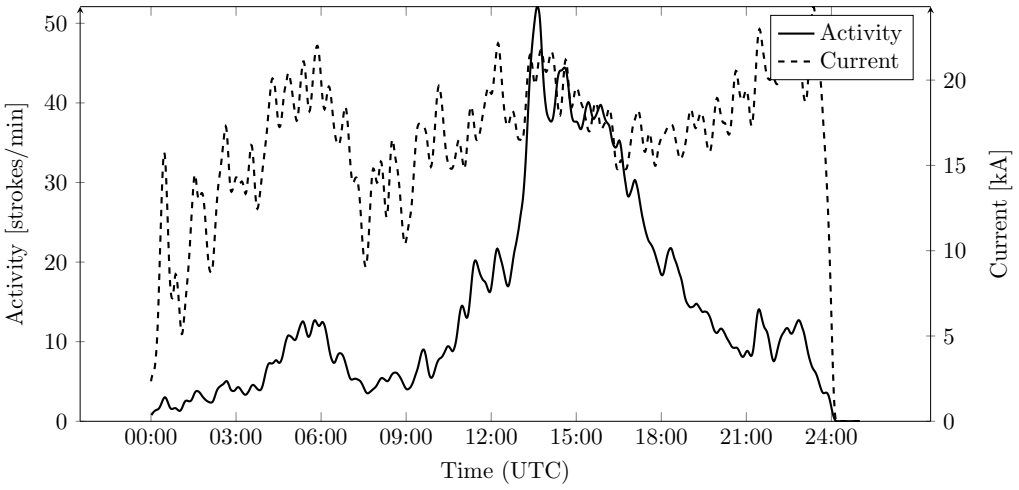
At local midday (10:00 UTC), when the amount of incident radiation from the sun peaks, cloud-to-ground lightning activity begins to increase. The cloud-to-ground process rapidly decays as the intracloud process peaks which is similar to the general temporal development presented in *Chapter 5*.

As the sun sets (18:00 UTC) there is a decrease in lightning activity and unlike the previous example there is not an increase in positive cloud-to-ground lightning. This is because the convective processes driven by latent surface energy are not as well established (cloud cover minimised the energy absorbed by the surface of the earth) and so a likely explanation is that the clouds were not as tall and that the height of the upper positive charge layer is lower.

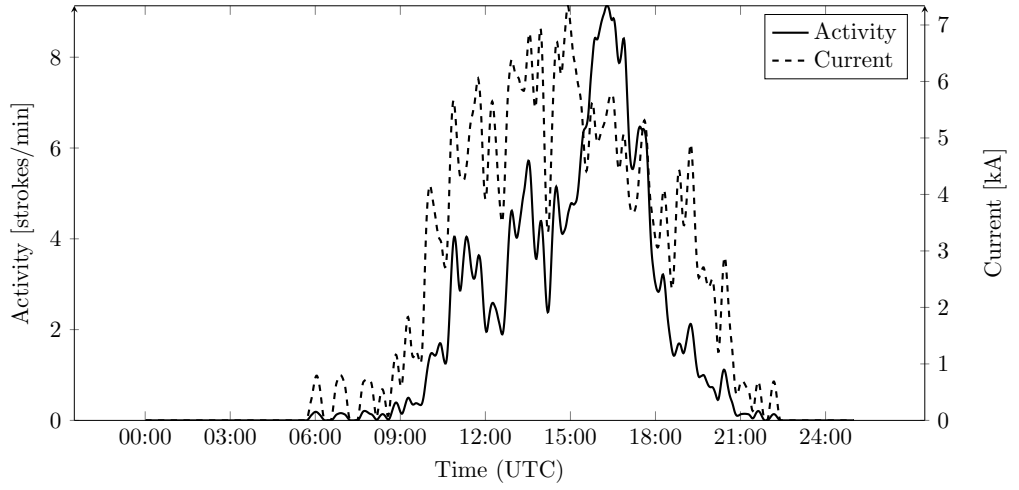
Figure 6.4 illustrates the likely cloud types and charge locations on 26 November 2007:

***Figure 6.4* - A:** The consistent layer of cloud across the figure indicates the presence of cloud screening the surface of the earth from the incident solar radiation. Solar radiation is at a maximum and energy within the cloud drives shallow convection, limiting cloud height. Charge separation, at depressed rates, occurs and cloud-to-ground lightning is observed.

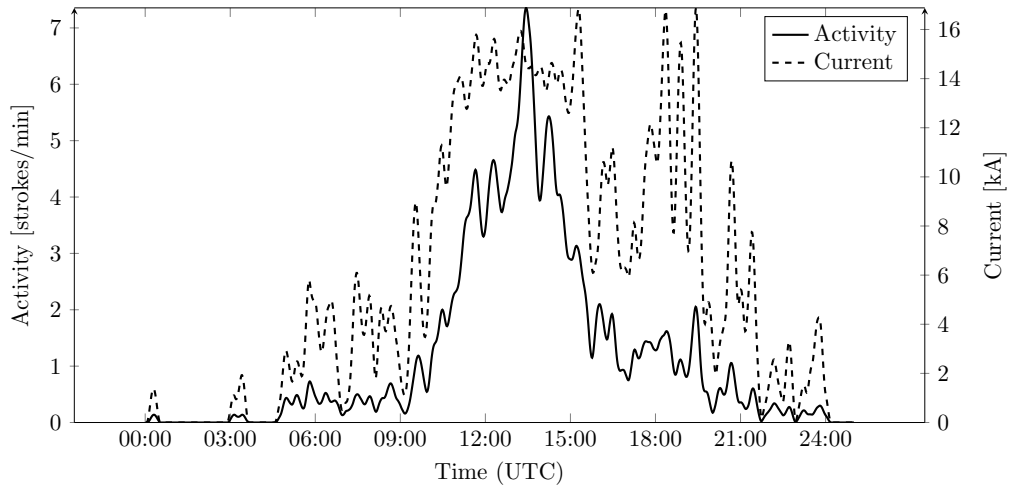
***Figure 6.4* - B:** Areas of sustained convection dissipate as the sun sets. The limited vertical extent of clouds prevented the formation of an isolated upper positive



(a) Negative cloud-to-ground strokes



(b) Positive intracloud strokes



(c) Positive cloud-to-ground strokes

Figure 6.3: Temporal development of activity and current observed at the modes, for the storms on 26 November 2007.

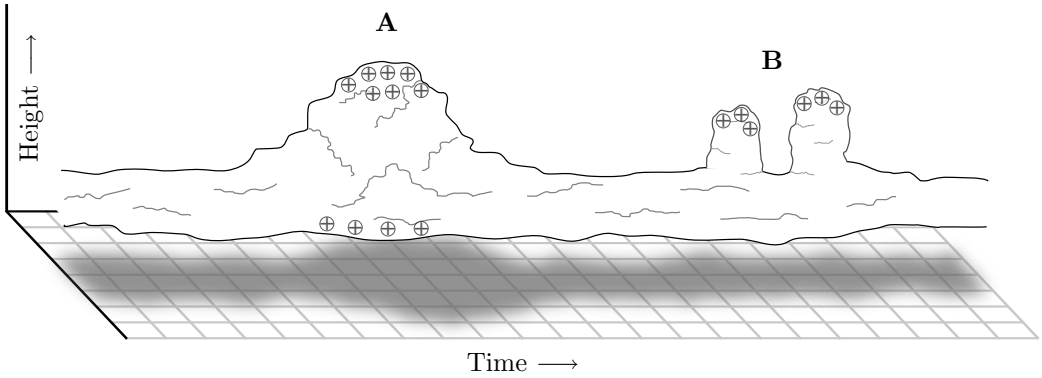


Figure 6.4: Illustration depicting the development of clouds and likely locations of positive charge on 26 November 2007.

charge layer and thus a second peak in positive cloud-to-ground activity is not observed.

6.4 19 February 2009

As shown in *Figure 6.6*, there is no lightning activity before 09:00 UTC on 19 February 2009, similar to the example of 17 November 2007. However, unlike on 17 November 2007, lightning activity begins to increase the rate of increase is relatively constant up to the peak in negative cloud-to-ground activity at 14:00 UTC. After the peak there is a five hour period where lightning activity is between 30 strokes/min and 50 strokes/min (compared to the peak of 63 strokes/min on 17 November 2007).

As negative cloud-to-ground activity peaks there is a dramatic drop in positive cloud-to-ground activity from a peak of 13 strokes/min to 2 strokes/min. The peak current for the positive cloud-to-ground mode is between 15.0 kA and 17.0 kA for the initial development of lightning activity, up to the end of the decay of positive cloud-to-ground at 16:30 UTC. This indicates that the predominant source of these positive cloud-to-ground strokes is the lower positive charge layer.

After the reduction in positive cloud-to-ground activity there is an increase in the peak current observed at the positive cloud-to-ground mode and values around 20.0 kA are observed (between 17:00 UTC and 18:00 UTC in *Figure 6.6*). This corresponds to the depletion of the upper positive charge layer. There is no noticeable increase in activity during this period because there is less convective energy and the vertical extent of the lightning producing clouds is limited, and so the upper positive charge layer is smaller.

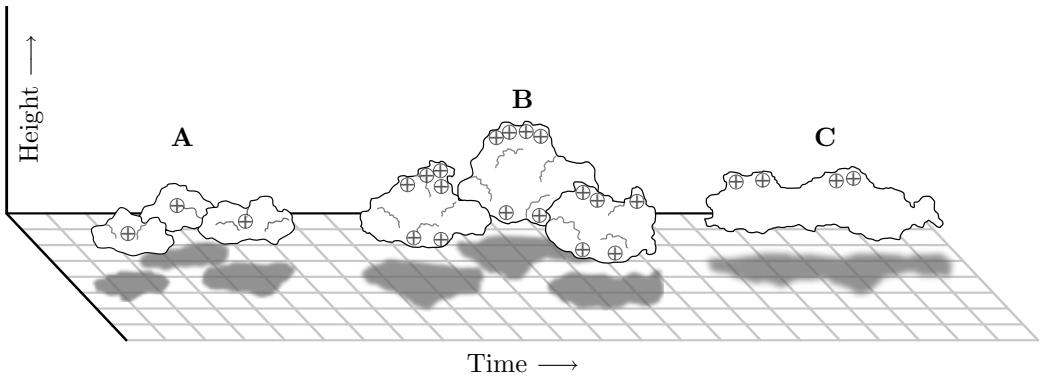


Figure 6.5: Illustration depicting the development of clouds and likely locations of positive charge on 19 February 2009.

***Figure 6.5* - A:** Likely initial cloud conditions.

***Figure 6.5* - B:** The negative cloud-to-ground lightning activity peaks when the convective processes are at a maximum and the rate of charge separation is also at a maximum.

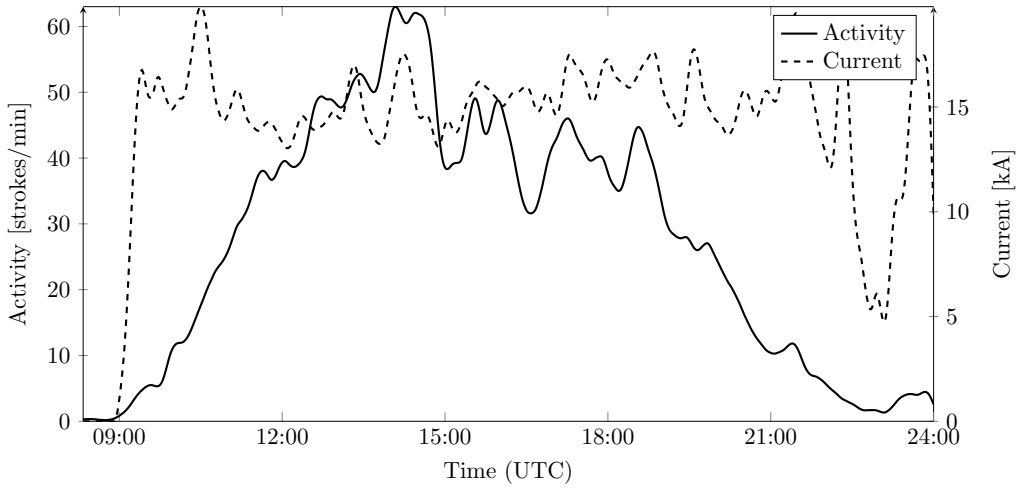
***Figure 6.5* - C:** The slowing down of the convection process leads to the collapse of the cloud structures.

Compared to the schematic in *Figure 6.2*, the vertical extent of the lightning producing clouds in *Figure 6.5* is much smaller.

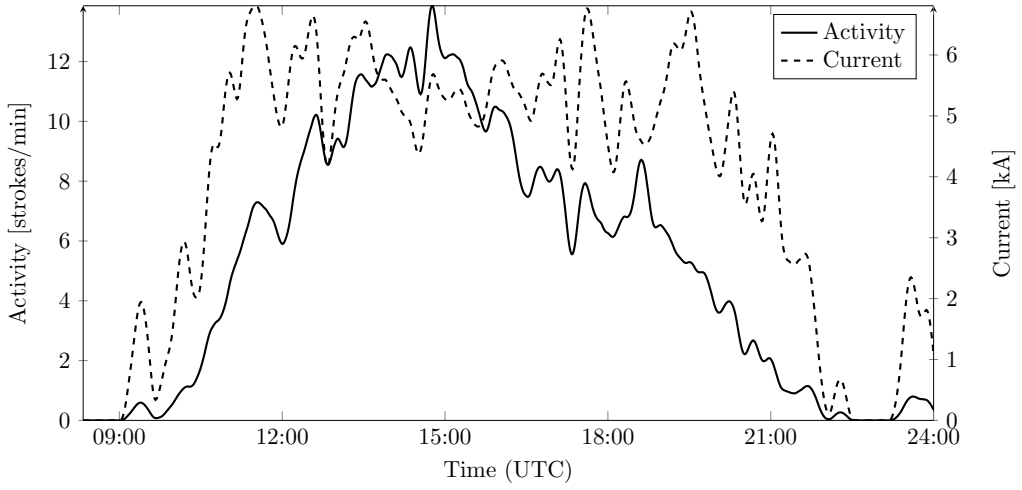
6.5 4 March 2010

Lightning activity on 4 of March 2010, shown in *Figure 6.8*, is interesting because the stroke rates are some of the highest observed by the SALDN since the commissioning of the network. The high level of lightning activity indicates the abundance of surface energy and high rates of energy transfer and thus there is sustained deep convection. The convection was enhanced by the presence of a large low pressure system over southern Africa.

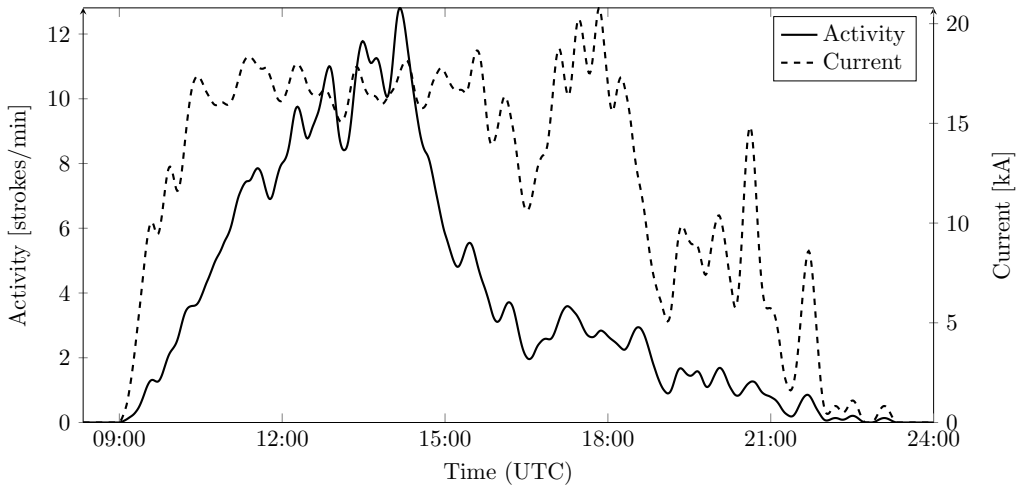
Before 09:00 UTC on 4 March 2010 there is no lightning activity, and after 09:00 UTC negative cloud-to-ground lightning activity increases over the next six hours to a peak of a 165 strokes/min—double that of the first example on 17 November 2007. The positive intracloud and cloud-to-ground activity measures are also higher on



(a) Negative cloud-to-ground strokes



(b) Positive intracloud strokes



(c) Positive cloud-to-ground strokes

Figure 6.6: Temporal development of activity and current observed at the modes, for the storms on 19 February 2009.

average than the preceding examples. The intracloud activity develops in line with the negative cloud-to-ground lightning activity, but is less peaked and the high levels are sustained for longer.

As negative cloud-to-ground and intracloud activity peaks at 15:00 UTC, there is a dramatic drop in the positive cloud-to-ground levels, which corresponds to the vertical development of the clouds and shielding of the upper layer of positive charge from ground by lower layers of charge.

As the negative cloud-to-ground stroke rate decreases, caused by reduced convection, there is an increase in positive cloud-to-ground strokes again. The lower layers of charge have been depleted through intracloud lightning and negative cloud-to-ground lightning and so the upper layer of positive charge is more successful in producing positive lightning to ground.

The illustration showing likely cloud development on 4 March 2010 is shown in *Figure 6.7*:

***Figure 6.7* - A:** Convection is assisted by the presence of a low pressure system over southern Africa and so vertical development of lightning producing clouds is quicker.

***Figure 6.7* - B:** The maximum vertical extent of lightning producing clouds is enhanced and so the upper layer where positive charge is located is larger and higher, potentially constrained by higher layers of the atmosphere.

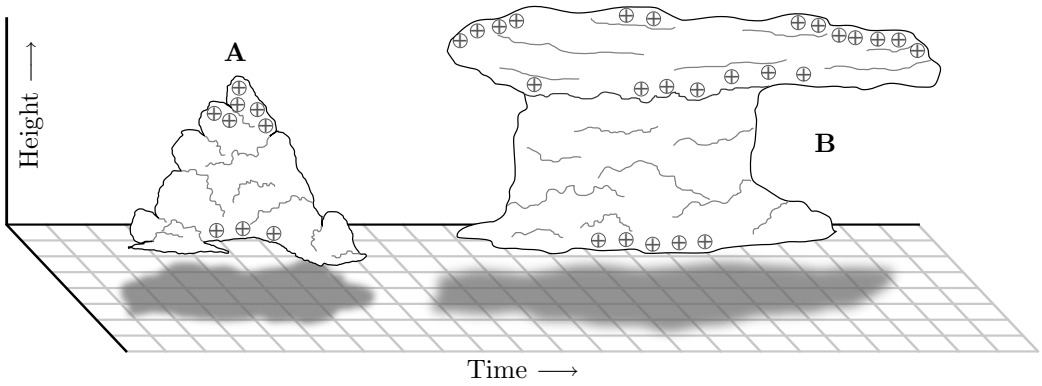
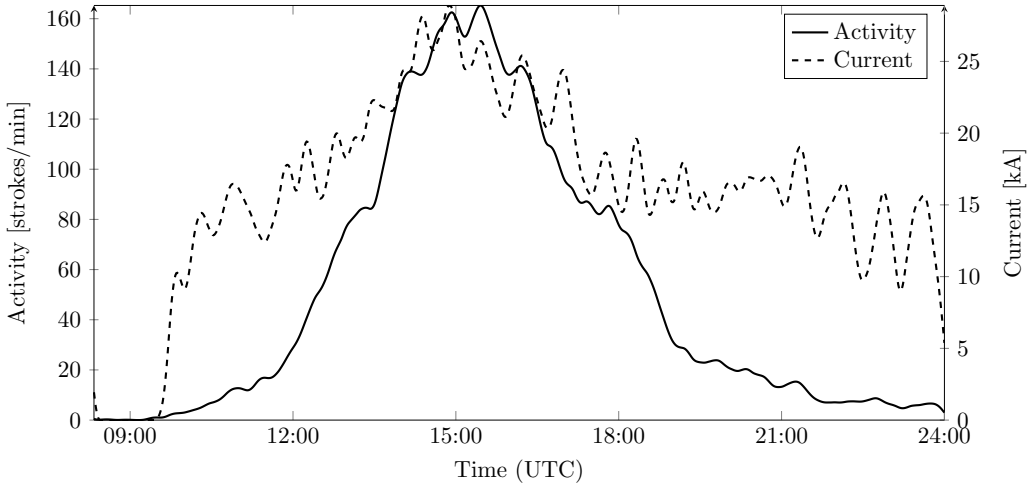
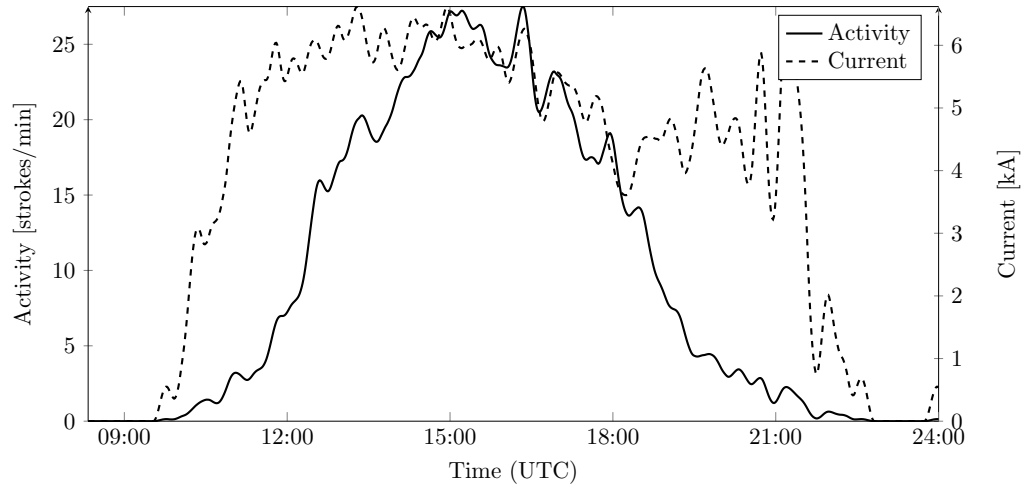


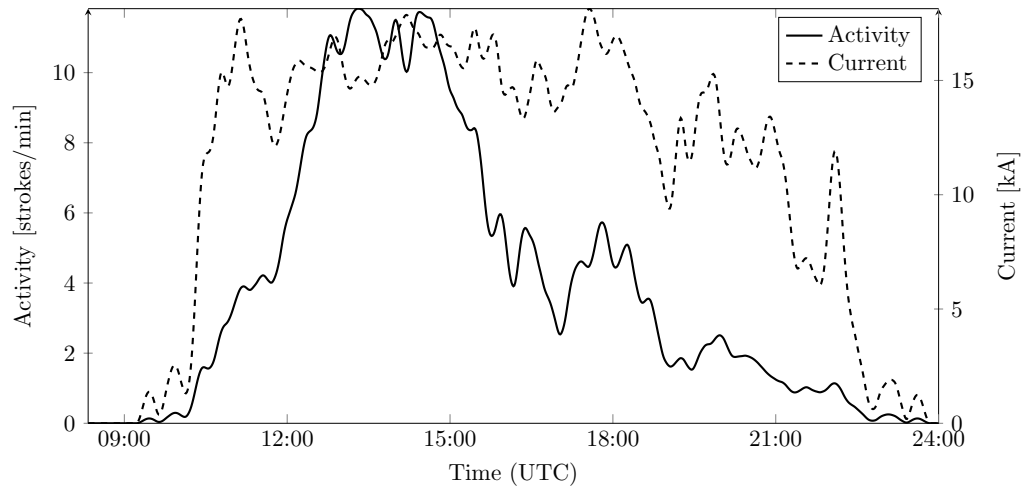
Figure 6.7: Illustration depicting the development of clouds and likely locations of positive charge on 4 March 2010.



(a) Negative cloud-to-ground strokes



(b) Positive intracloud strokes



(c) Positive cloud-to-ground strokes

Figure 6.8: Temporal development of activity and current observed at the modes, for the storms on 4 March 2010.

Summary

It was shown in this chapter that it is possible to measure the independent development of intracloud and cloud-to-ground components on as fine a time scale as a day. This self-consistent analysis accounts for the presence of misclassified strokes in the data sets and the method presented in *Chapter 4* and *Chapter 5* may be applied to a single day.

Four days were selected from the 801 day data set of the South African Lightning Detection Network. The lightning development for each day was presented as measures of the negative cloud-to-ground mode, and positive intracloud and cloud-to-ground modes. The shielding effect, where upper layers of positive charge are screened by lower layers of charge from producing lightning that attaches to ground, is illustrated in the examples. The development of clouds on the particular days is also given and the likely locations of positive charge for the upper and lower layers is illustrated.

The following chapter extends the modal measurement down to an isolated storm which is selected from a lightning day through a clustering algorithm.

Chapter 7

Temporal development of a select storm

This chapter presents a storm tracking algorithm that is used to select strokes that were part of a single storm. This subset of lightning strokes is then processed with the minute of the day algorithm introduced in previous chapters. The contribution of this chapter is in showing the spatial development of a storm and the corresponding lightning activity development. It is shown that the modal analysis method, developed in this thesis, is a valuable technique to be applied to single storms.

7.1 Storm tracking and analysis algorithm

The algorithm implemented and used to select strokes produced by the same storm system is similar to the cell tracking algorithm described by Lojou, J-Y. & Cummins (2006), but includes the ability for cells to merge. The algorithm is schematically illustrated in *Figure 7.1*.

The algorithm operates on a sequentially ordered list of strokes, and the location of each storm centre is computed as the centre of mass of the strokes assigned to that storm. The mass contribution of each stroke added to a storm is inversely proportional to the distance of the stroke from the storm centre when the stroke is added.

Only recent strokes are considered when computing the centre of mass, and the algorithm is stepped along in time by processing a ticker which is incremented by the latest stroke time. This provides an efficient way to process periods where there

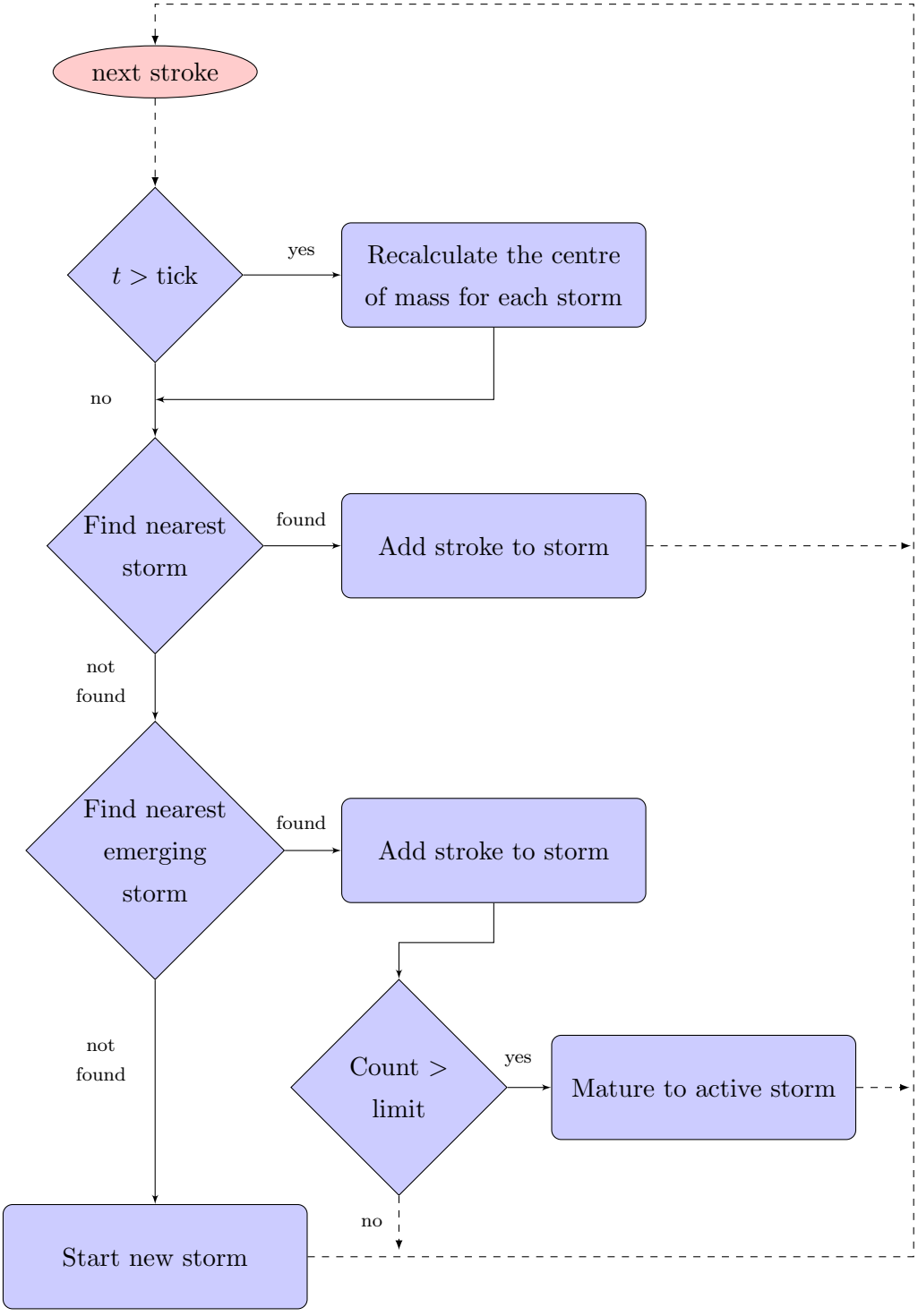


Figure 7.1: Algorithm for the assignment of a stroke to a storm, for a sequential time series of strokes.

is no activity without having to recompute storm parameters when no strokes were detected.

The location of a stroke is compared against the locations of active storms and is added to the closest storm. However, if no active storm is found the stroke is then compared against a list of emerging storms. The emerging storm assignment is a temporary classification assigned to a stroke when it is not near any active or emerging storm, and consists of a unique identifier which is assigned to all other strokes that are part of that storm. Once a sufficient number of strokes have been assigned to an emerging storm, it is matured into an active storm. This process prevents spurious 'bolt out of the blue' events from forming a storm that needs to be tracked.

When the active region of two storms overlap, the storms coalesce and strokes occurring after the merger are assigned to the oldest of the two storms. The resulting active region of the merged storms is the union of the two areas, and the centre of mass is recomputed off the combined active region.

When a stroke is compared against each storm or emerging storm, the time difference between the last stroke added to a storm and the current stroke is computed and if it exceeds a no activity duration constant, the storm is the expired and removed from the lists.

All the strokes from a storm may be selected by querying the storm identifier that was assigned to each stroke by the algorithm. The minute of the day algorithm, as described previously in *Chapter 5*, is then run on the subset of strokes belonging to that storm.

7.2 Spatial development of the storm

The storm begins in the east of southern Africa at 14:00 UTC on 17 November 2007 in the region south of Swaziland and east of Lesotho. The region is highlighted in *Figure 7.2*. The storm passes over Durban, a commercial city and port in South Africa. The storm selected for analysis from the set of 64 million strokes is interesting for a variety of reasons:

- lightning activity begins as isolated cells producing lightning;
- these separate cells coalesce over land; and

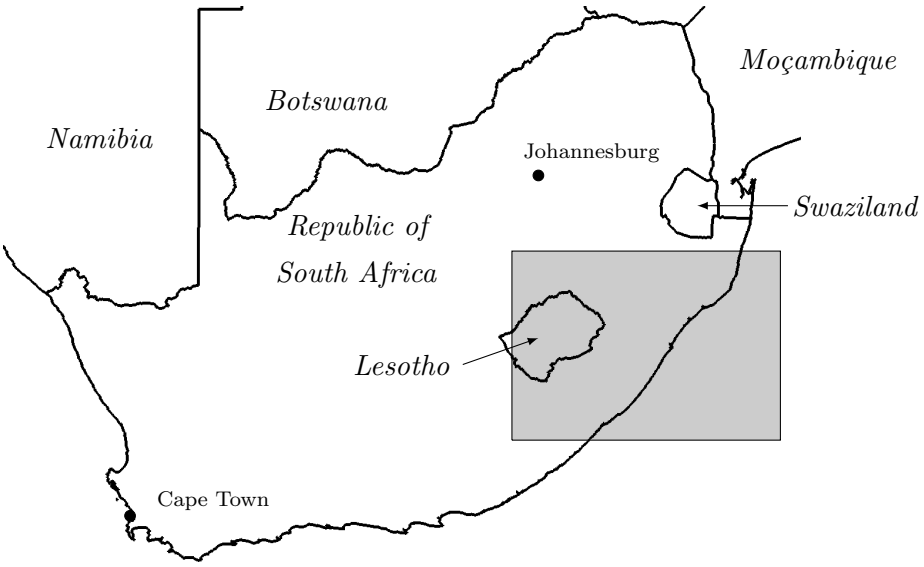


Figure 7.2: Map of southern Africa, showing the context of storm area for the time series in *Figure 7.3*.

- the large storm, originating from the independent cells, moves over the Indian ocean where it dissipates.

The panels shown in *Figure 7.3* are 10 minute snapshots of lightning activity spaced 30 minutes apart. The first panel is as the cells develop, and several sparse areas can be seen to the east and south east of Lesotho. In the next three panels, which presents the development of the cells over an hour and half, show the cells moving east and the northern groups moving south to coalesce with the larger body. This body continues to move east—offshore. The final four panels show the storm broadening and stroke activity reducing.

The fact that the cells coalesce over land is interesting because there are not any orographic features that caused the cells to combine and so the forces responsible for the combination are meteorological. The dissipation over the ocean is important because the surface is roughly homogeneous and there is less convective energy available to support tall cumulonimbus thunder cells and hence the convection process responsible for charge separation collapses and lightning activity ceases.

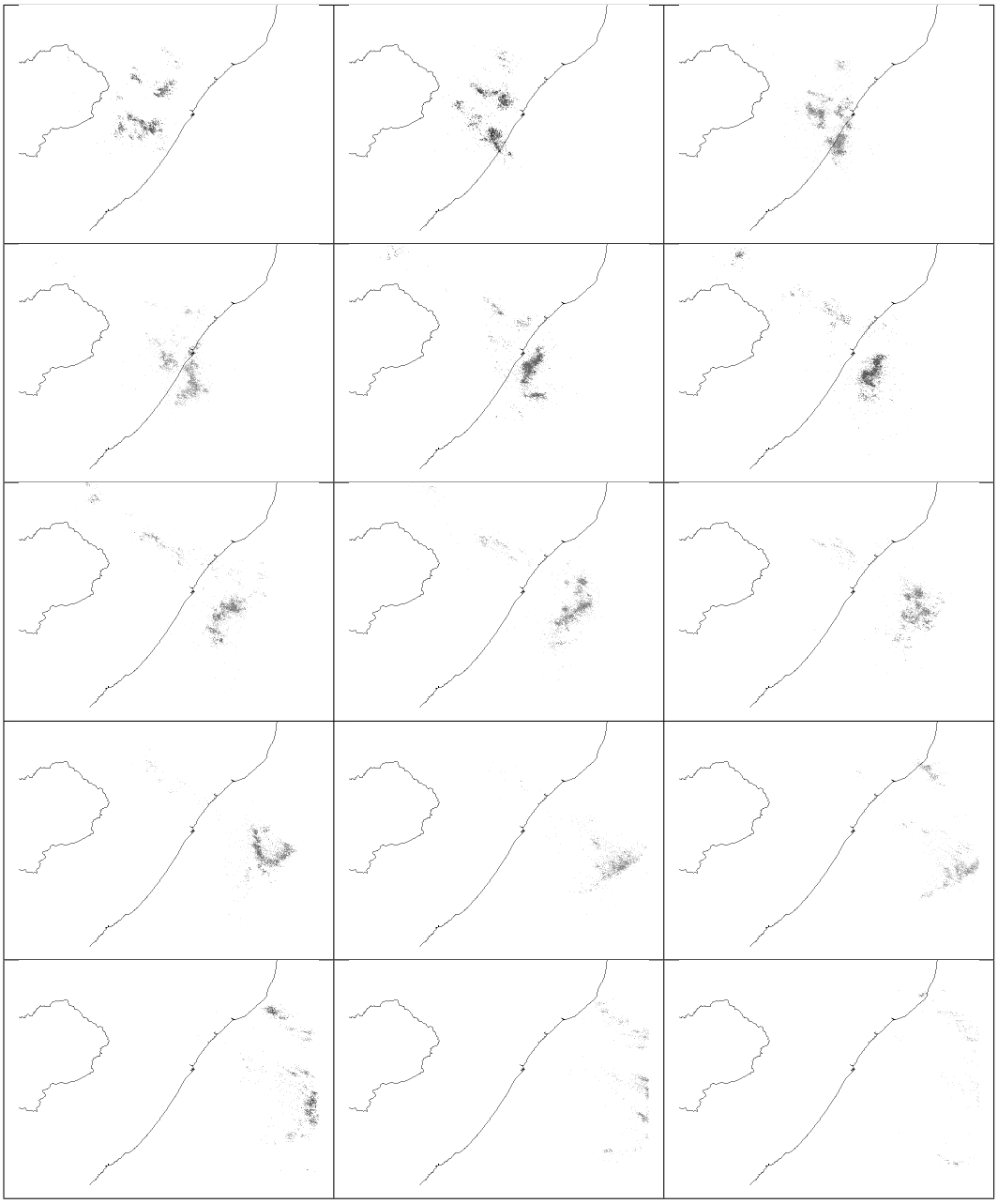


Figure 7.3: Panels showing the development of a storm, as the set of recorded strokes, shown in panels spaced 30 minutes apart. The time series starts in the top left corner, runs left to right and then top to bottom. The border of Lesotho and the coast line are also visible. The storm starts inland and moves over the sea where it dissipates.

7.3 Lightning development of the storm

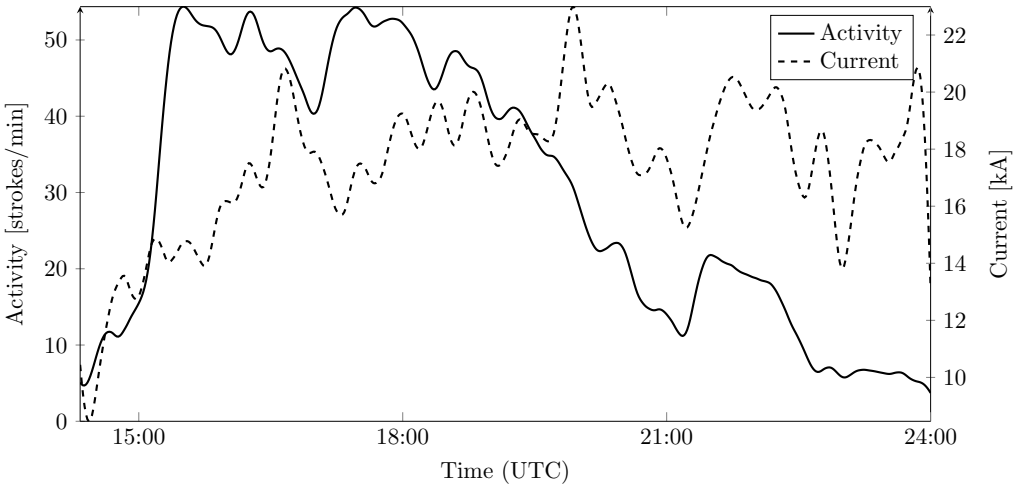
The three measures of lightning activity, negative cloud-to-ground, positive intracloud, and positive cloud-to-ground, develop simultaneously as the cells start around 14:00 UTC. The reduction in activity in all three measures corresponds to the off shore movement of the storm, as shown in the second row of *Figure 7.3*. This event corresponds to the start of final stage of the storm, which ends at the end of the day.

The negative cloud-to-ground stroke rate is as high as 55 strokes/min, which corresponds to a peak flash rate of 20 strokes/min if a flash multiplicity of 2.5, from Gill (2008), is used. For comparison, mean flash rates as high as 60 flashes/min have been recorded in previous independent studies (Rakov & Uman (2003)).

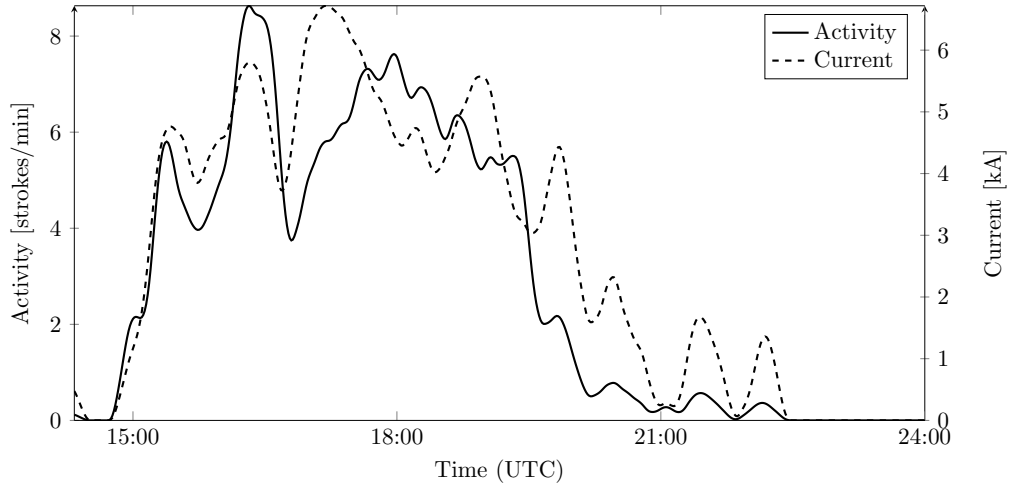
After the initial development of the storm over land, the positive cloud-to-ground component consists of three distinct periods of activity, each at a lower level than the preceding period. The first period is between 17:00 UTC and 19:UTC and corresponds to the dissipation of the outlying cells over land. As the primary cell moves offshore there are sparse centres of lightning activity that separate from the main cell and move north where they dissipate.

The second period of activity, between 20:00 UTC and 21:00 UTC corresponds to the breakup of the storm mass over the sea and the dissipation of outer extent of the storm mass. There is also a rapid decrease in the positive intracloud component, which indicates less energy from convection—this intuitive conclusion is derived from the fact that the storm is no longer over land and the convective processes over the sea are depressed.

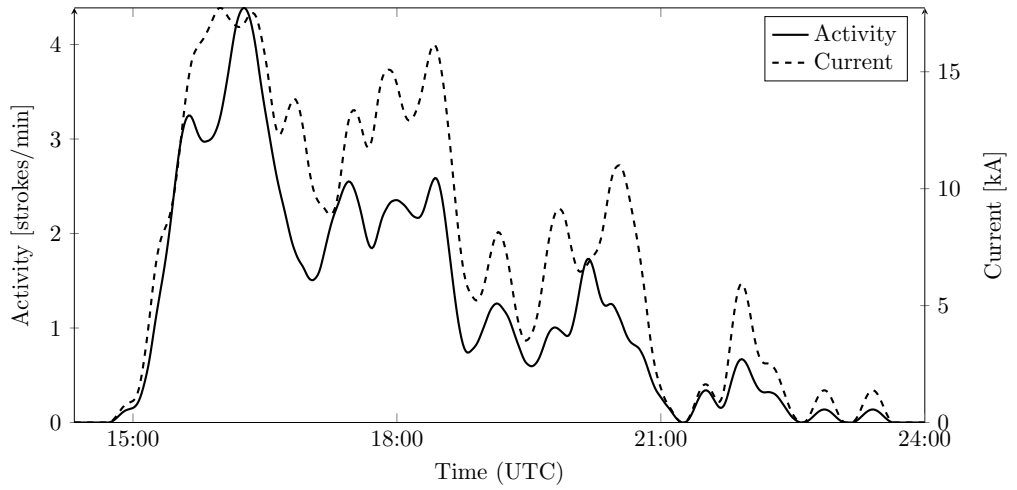
The final period of positive cloud-to-ground activity starts at 21:30 UTC and corresponds to the final dissipation of the storm that started 9 hours before. There are few positive cloud-to-ground strokes as the remaining charge centres are neutralised. There is also a short increase in the negative cloud-to-ground activity which occurs because the convective process has stopped and in addition to the upper layers of positive charge, the upper layers of negative charge are also discharging to ground.



(a) Negative cloud-to-ground strokes



(b) Positive intracloud strokes



(c) Positive cloud-to-ground strokes

Figure 7.4: Temporal development of activity and current observed at the modes, for a storm on the 21 November 2007.

Summary

This chapter concludes the modal analysis and shows that the application of the modal technique provides a measure on as fine a time scale as a storm. To illustrate this fine time resolution, a single storm was analysed. The storm begins with separate cells converging over land and then the central storm mass moves offshore where it dissipates over the ocean.

The spatial development shows that the cells coalesce because of meteorological phenomena and not orographic confinement. Lightning activity in the storm decreases as it moves over the ocean. It is inferred that the convective process is depressed because energy transfer from the ocean mass to the surface air to drive convection is less than the same processes over continental masses. The decay of storm activity is characterised by three periods of increased positive cloud-to-ground lightning, each of which correspond to the dissipation of a component of the storm.

The next chapter presents the variation of lightning stroke parameters with altitude over a section of southern Africa. The section of southern Africa used for the analysis has severe topography and presents a unique opportunity to investigate the variation of lightning stroke parameters across altitudes in a spatially confined area.

Chapter 8

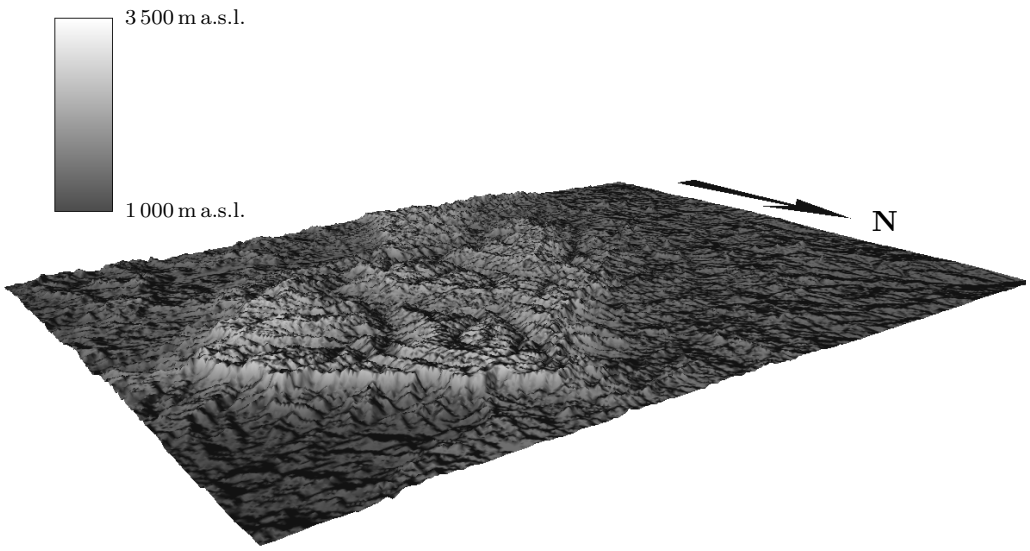
Orographic effects

The previous chapters focused on the temporal variation of lightning activity, and presented a novel method for measuring the variation of separate processes within the same data set. This chapter looks at the orographic variation of stroke parameters. The contribution of this chapter is in showing a specific altitude sensitivity in cloud-to-ground lightning stroke densities below 2150 m above sea level (m a.s.l.) and for negative intracloud stroke densities above 2 150 m a.s.l.

8.1 The Drakensberg and Lesotho

The Drakensberg mountain range is found in the east of the Republic of South Africa, and the watershed on the northern extent is the boarder between the enclave of Lesotho and the Republic of South Africa. The topography of the Drakensberg is severe and is characterised by a wall of rock between 500 m and 800 m high. A section on the northern extent, which is a few hundred kilometres long, has such a severe wall of rock that it is called the Amphitheatre. *Figure 8.1a* is a rendered view looking at the amphitheatre from the north east, with no vertical exaggeration.

The highest point in the Republic of South Africa and Kingdom of Lesotho is found near the northern tip of the Drakensberg, and a high plateau in this region is visible in *Figure 8.1b*. The altitude of the low-lands around the Drakensberg is roughly 1 400 m a.s.l. and the Indian Ocean is to the east and south east. There are no mountains or other contributing orographic features to the north, west or south of the range. The combination of location and severe topography provides a unique arrangement that may be used to study the variation in stroke density with altitude.



(a) Rendered 3D view

(b) Planar view

Figure 8.1: Rendered view of the Drakensberg, looking south. Shading is proportional to altitude ([m] above sea level), with the highest point some 3 300 m a.s.l., and the general area some 1 000 m a.s.l.

8.2 Measurement

The stroke density with altitude was calculated with the same 801 day data set that was used in the previous diurnal and temporal analyses. Stroke counts were recorded on a 50 m × 50 m grid and then interpolated. The cumulative stroke count for the interpolated data set was normalised so that the cumulative stroke count is equal to the number of strokes in the data set before it was interpolated.

A mask was then created from the digital elevation model, where all areas above a specific altitude were selected and only strokes within the mask were counted. A true surface area, discussed below, was also computed for the masked area. The per contour stroke count and density was calculated as the differential of the cumulative stroke count through altitude.

Thus stroke density at a specific altitude may be calculated as:

$$S_d = \frac{\text{Count of strokes in contour level}}{\text{Contour area}} \tag{8.1}$$

8.2.1 Slope corrected surface area

As the stepped leader approaches earth, the ground based electric field is not only a function of the amount of charge at the head of the leader but also of the ground geometry. Thus stroke density must be a function of real, or a close approximation to, surface area. Stroke density should not be a function of a two dimensional surface area. As illustrated in *Figure 8.2*, Geographic Information Systems (GIS) inherently process two dimensional cells that only represent the planar surface area. It is possible to correctly represent the area by calculating the slope from a digital terrain model.

The slope corrected surface area is calculated from the slope and planar model according to the method described by Grohmann (2004):

$$h = b \tan\theta \tag{8.2}$$

$$b' = \sqrt{b^2 + (b \tan\theta)^2} \tag{8.3}$$

$$\text{Area} = b' \cdot l \tag{8.4}$$

where

l = Grid length [m]

b = Grid breadth [m]

h = Slope height [m]

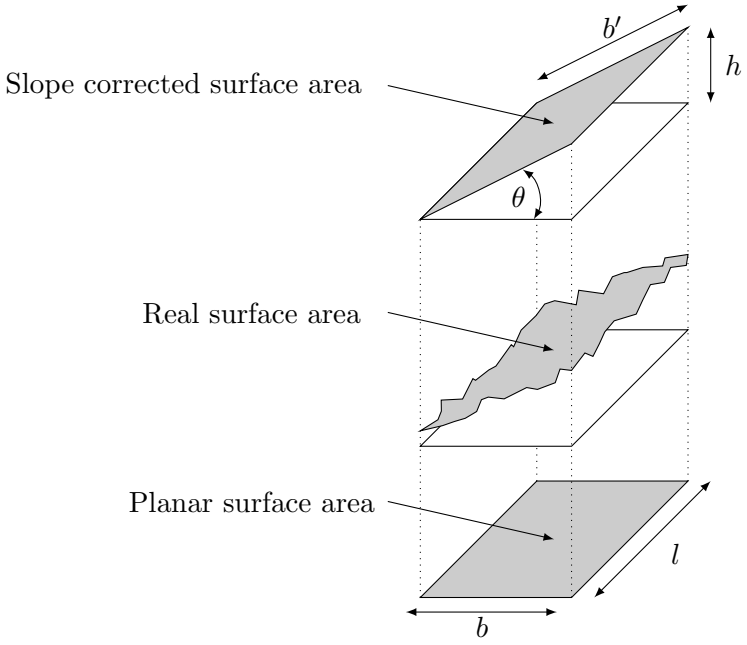
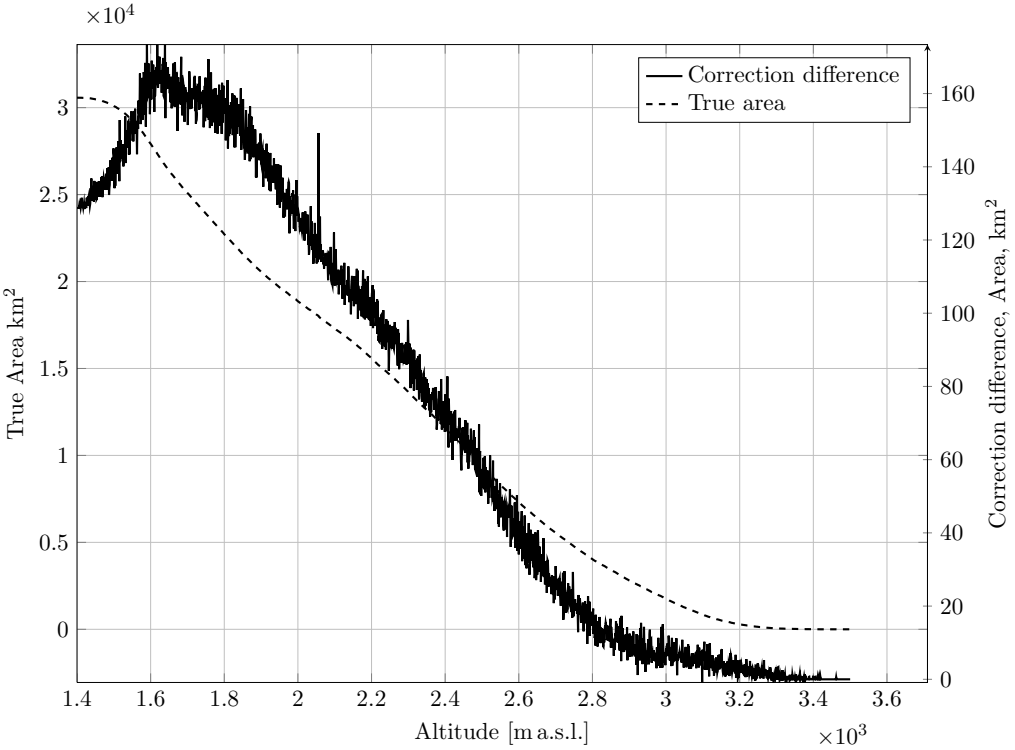


Figure 8.2: Illustration of real surface area, planar surface area, and slope corrected surface area, adapted from Grohmann (2004)

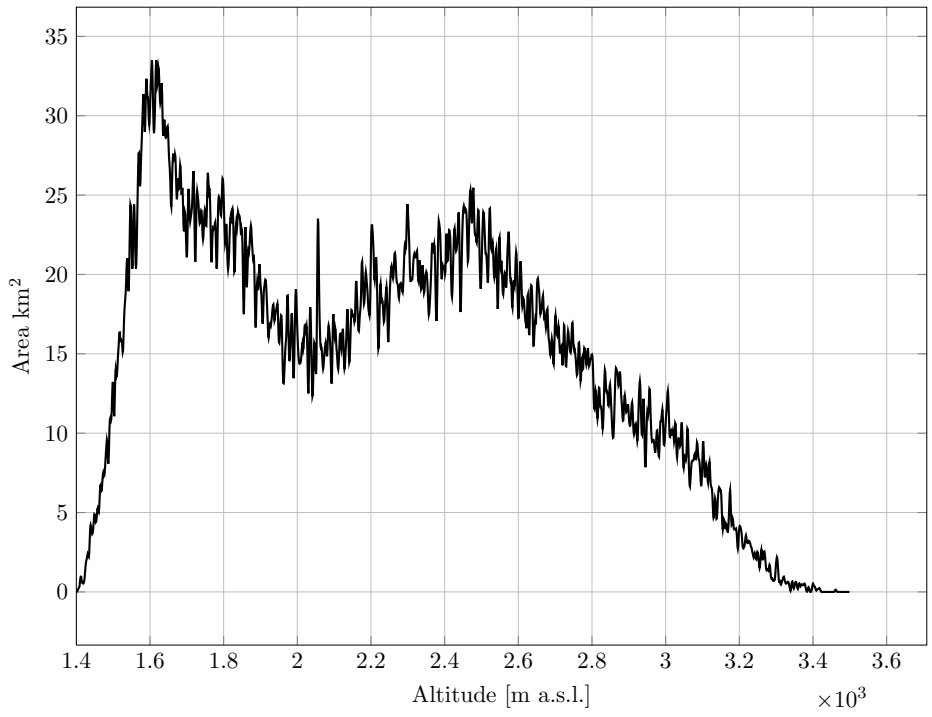
This equation will not calculate the slope corrected area for a slope where there is an inclination along l (i.e. $l' > l$); and so to correct for this a regular grid where $b = l$ must be used.

Figure 8.3a shows the slope corrected surface area and the difference between the planar surface area and slope corrected surface area. Given the magnitude of the maximum correction applied to the surface area, this correction is required for the severe topography of the Drakensberg and Lesotho when calculating area based parameters.

Figure 8.3b shows the resulting area per 1 m contour for the Drakensberg and Lesotho region, starting at 1 400 m a.s.l. and ending at 3 400 m a.s.l. The area at 1 400 m a.s.l. is relatively small because of the confinement of the mask, but the corresponding count is less and hence the stroke density is unaffected. The contour with the greatest surface area is found at 1 600 m a.s.l. which is the top of a sedimentary sand stone layer that underlies the volcanic basalt of the Drakensberg. The second peak at 2 500 m a.s.l. corresponds to the large section of slopes to the south of the mountain range, and then the area decreases almost linearly to just below the highest point at 3 400 m a.s.l.

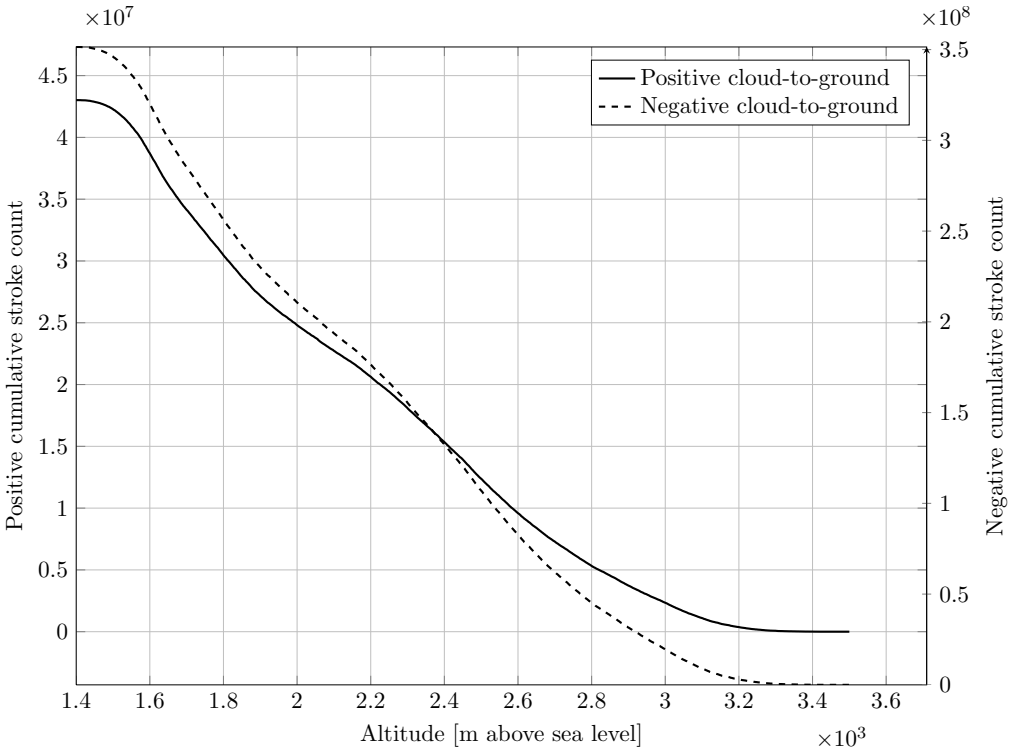


(a) Slope corrected surface area, and the area difference between the planar area and corrected area, as a function of altitude

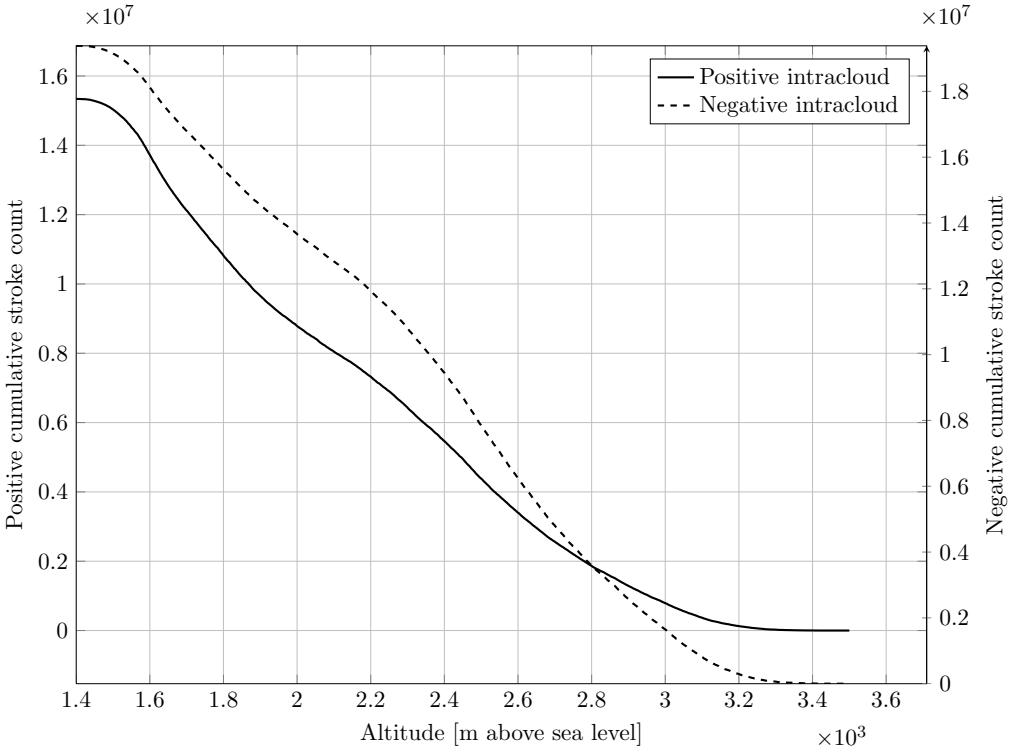


(b) Area per contour as a function of altitude

Figure 8.3: Surface area and contour area for the Drakensberg and Lesotho region.

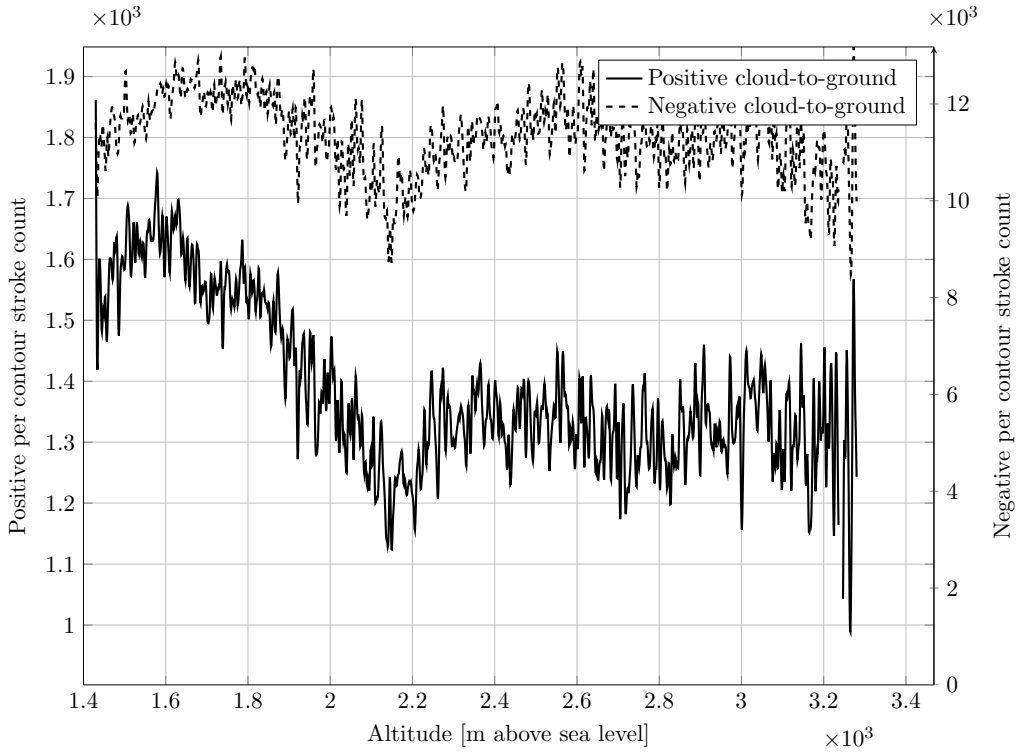


(a) Cumulative stroke count for cloud-to-ground strokes

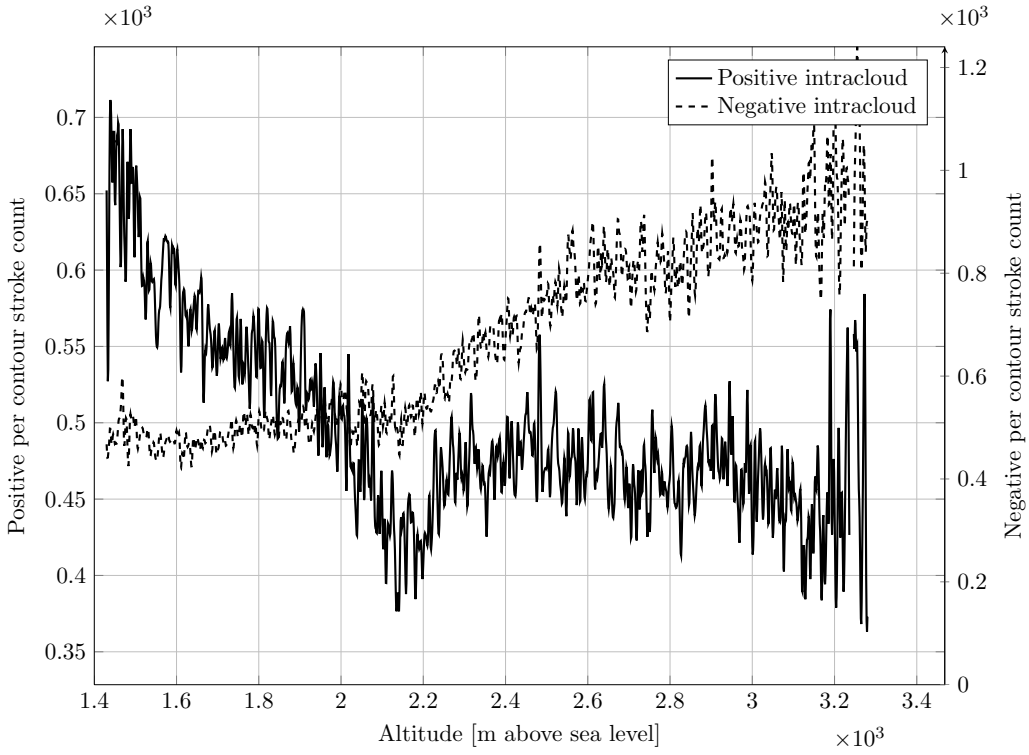


(b) Cumulative stroke count for intracloud strokes

Figure 8.4: Cumulative stroke count as a function of altitude for the Drakensberg Lesotho region.



(a) Per contour stroke count for cloud-to-ground strokes



(b) Per contour stroke count for intracloud strokes

Figure 8.5: Per contour stroke count as a function of altitude for the Drakensberg Lesotho region.

8.3 Stroke count, density and altitude

The cumulative stroke count as function of altitude, which is the sum of all strokes occurring at and above a specific altitude for cloud-to-ground strokes is shown in *Figure 8.4a* and for intracloud strokes in *Figure 8.4b*. The cumulative count for negative cloud-to-ground strokes, shown in *Figure 8.4a* is an order of magnitude larger than the cumulative count for the positive cloud-to-ground strokes: the lightning detection network classified 52.7 million strokes as negative cloud-to-ground and 7.2 million strokes as positive cloud-to-ground (without any reclassification limits applied). The shape of the cumulative profiles for negative and positive cloud-to-ground strokes are similar and little may be inferred from the cumulative profiles.

Figure 8.4b shows the cumulative stroke count for intracloud strokes. The positive intracloud cumulative altitude profile is similar to the cloud-to-ground cumulative altitude profiles, but the negative intracloud cumulative profile is different from 2 500 m a.s.l. where the decay is almost linear.

The stroke count per contour area, as defined by *Equation 8.1*, for cloud-to-ground strokes is shown in *Figure 8.5a* and for intracloud strokes in *Figure 8.5b*.

As shown in *Figure 8.5a*, the cloud-to-ground data sets exhibit an inverse relationship between altitude and stroke count for altitudes between 1 600 m a.s.l. and 2 150 m a.s.l. At 2 150 m a.s.l. the stroke count is at a minimum, and above this altitude the positive cloud-to-ground stroke count is insensitive to altitude. Above 2 150 m a.s.l. the negative cloud-to-ground stroke count increases with altitude and then above 2 600 m a.s.l. the stroke count is invariant.

The positive intracloud stroke count as a function of altitude, shown in *Figure 8.5b*, is similar to the cloud-to-ground data sets: and stroke count decreases with altitude to a minimum at 2 150 m a.s.l. and then increases with altitude up to 2 400 m a.s.l. and then decreases at a reduced rate with altitude.

The negative intracloud data set, shown in *Figure 8.5b*, is not insensitive to altitude above 2 150 m a.s.l. The negative intracloud stroke count increases with altitude from 1 400 m a.s.l. to 3 300 m a.s.l. This consistent trend is not present in any of the other data sets. Nor has the altitude sensitivity of lightning strokes with assigned intracloud type classification been documented elsewhere.

Figure 8.6 shows the spatial distribution of cloud-to-ground strokes and *Figure 8.7* shows the spatial distribution of intracloud strokes; rendered views are shown in

Figure 8.8 and *Figure 8.9* respectively.

The shape and distribution of valleys, across altitudes less than 2 150 m a.s.l., are visible in *Figure 8.6a*. Similar features are discernible in *Figure 8.6b*, but with less spatial resolution since there are fewer positive cloud-to-ground strokes.

The linear relationship of negative intracloud strokes with altitude is visible in *Figure 8.7a*, with higher strokes counts on peaks above 2 150 m a.s.l. The positive intracloud spatial distribution, shown in *Figure 8.7b* does not show the same sensitivity to altitude and high stroke counts are seen in regions to the the north east of the severe topography.

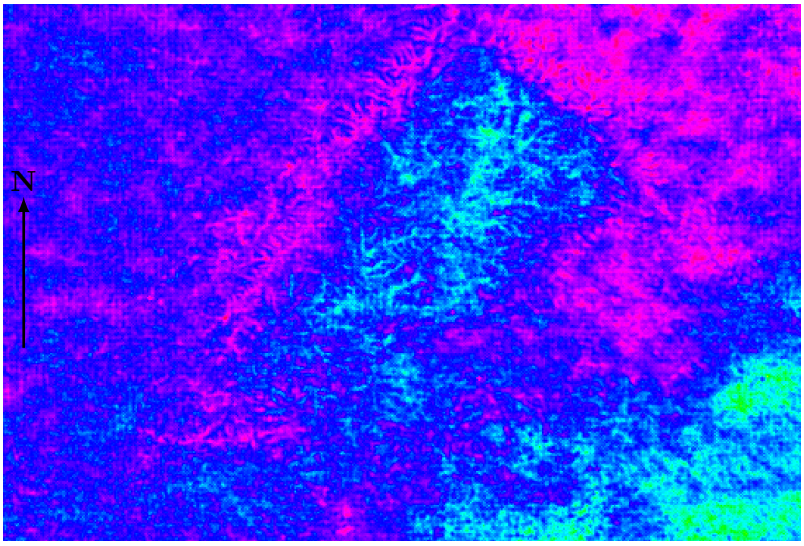
Summary

This chapter concludes the analysis of the orographic variation of lightning stroke density and the analysis of lightning stroke data over southern Africa.

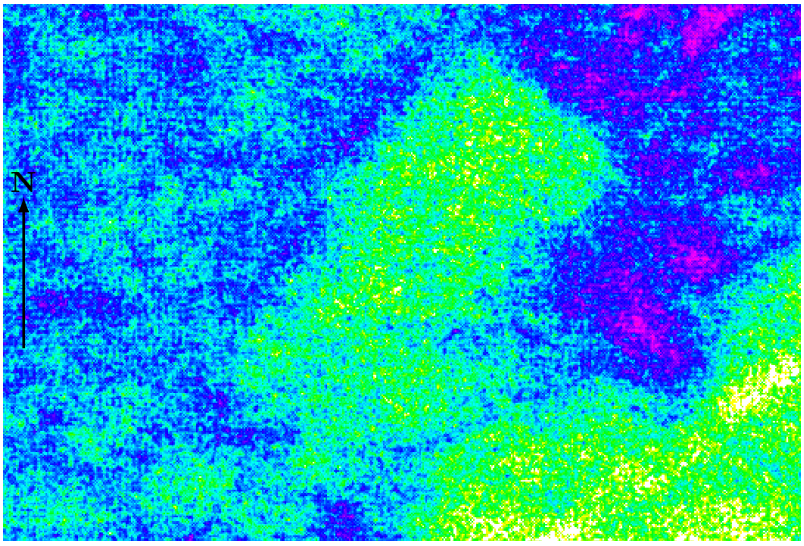
Lightning stroke density was calculated across the altitudes of the Drakensberg, using slope corrected surface area. The analysis was computed between 1 400 m a.s.l. and 3 500 m a.s.l. with 1 m vertical resolution.

It was shown in this chapter that negative intracloud stroke density increases linearly with altitude above 2 150 m a.s.l. Negative and positive cloud-to-ground lightning stroke density varies inversely with altitude between 1 400 m a.s.l. and 2 150 m a.s.l.

The subsequent chapter presents the conclusion of the thesis.



(a) Negative

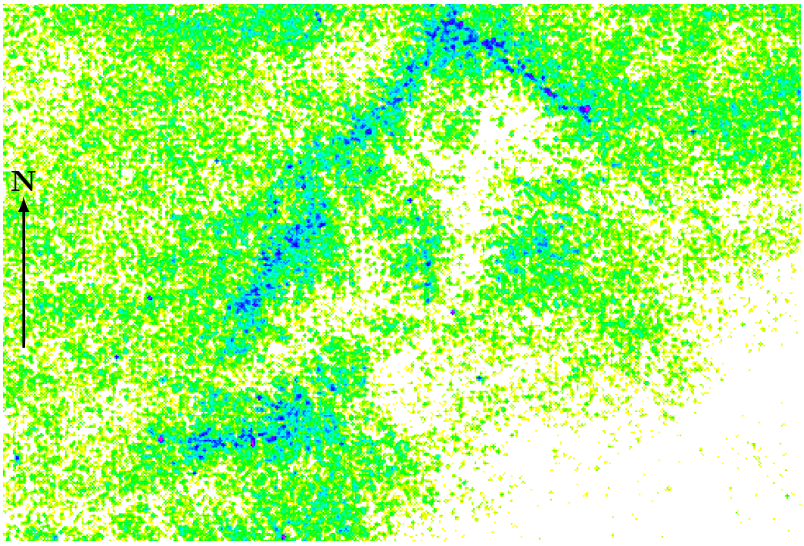


(b) Positive

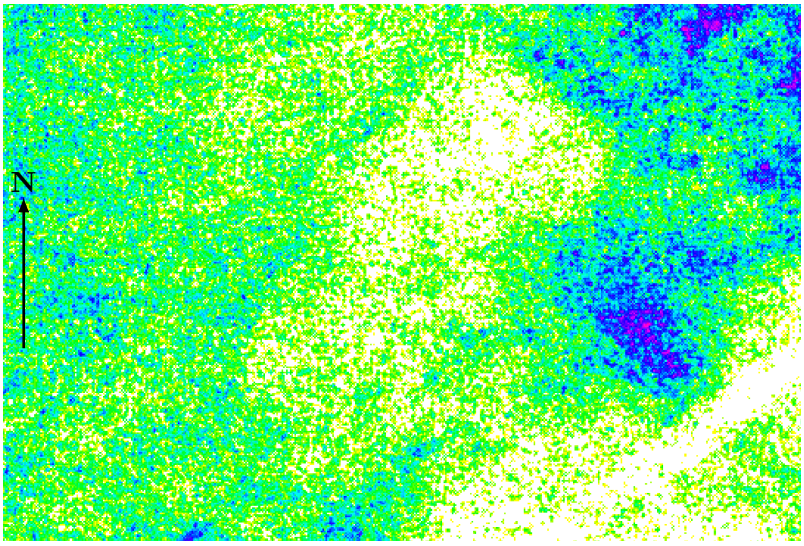
0 (no activity)  1 (highest activity)

(c) Normalised Scale for above figures

Figure 8.6: Planar view of the spatial distribution of cloud-to-ground strokes over the Drakensberg region. Colouring is proportional to a normalised stroke count, the elevation model and view is shown in *Figure 8.1*.



(a) Negative

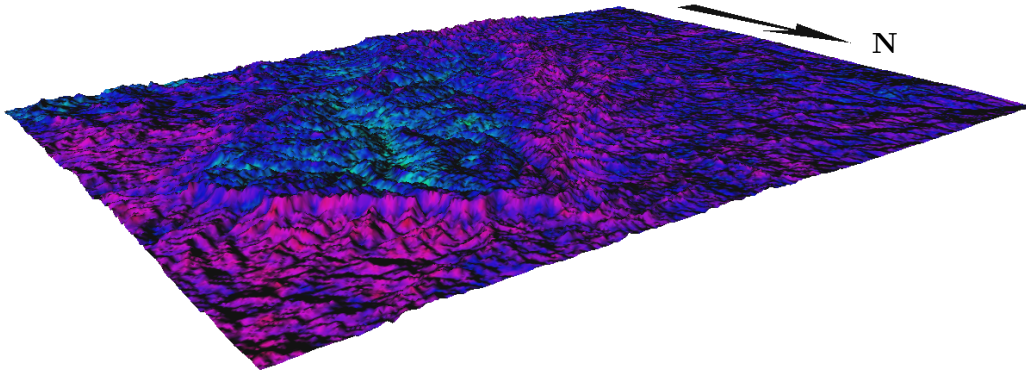


(b) Positive

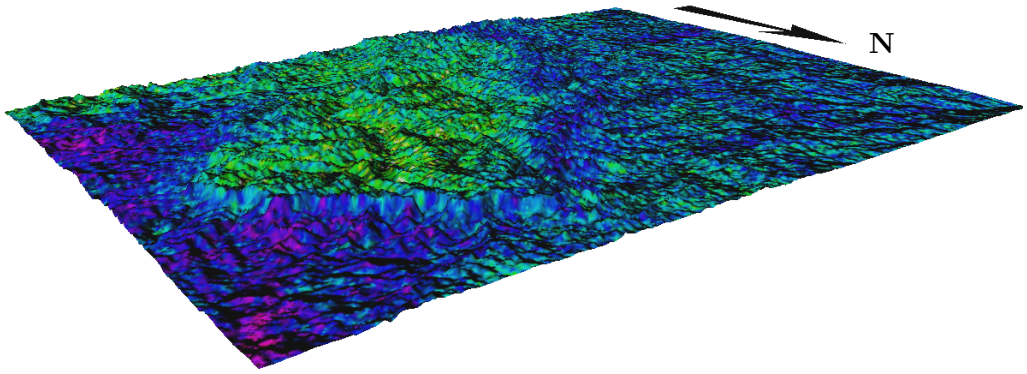
0 (no activity)  1 (highest activity)

(c) Normalised Scale for above figures

Figure 8.7: Planar view of the spatial distribution of intracloud strokes over the Drakensberg region. Colouring is proportional to a normalised stroke count, the elevation model and view is shown in *Figure 8.1*.



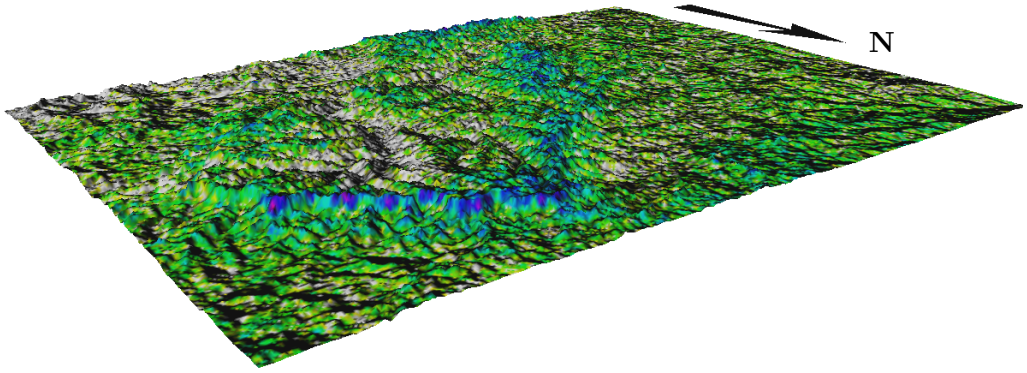
(a) Negative



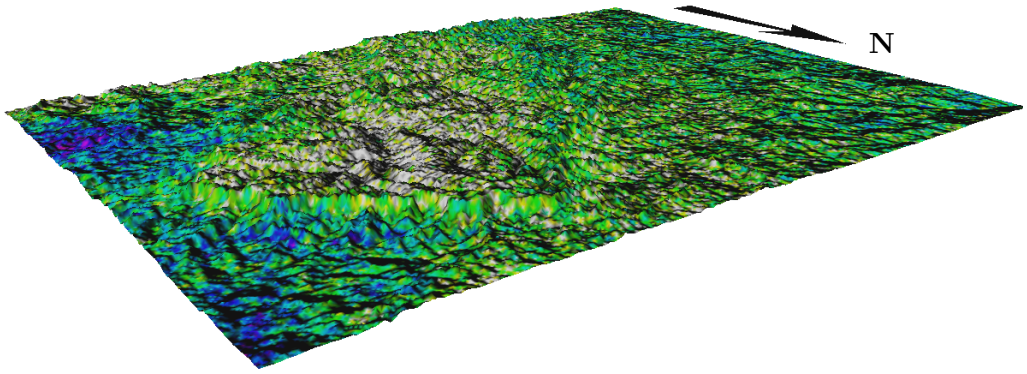
(b) Positive

(c) Normalised Scale for above figures

Figure 8.8: Rendered view of cloud-to-ground strokes over the Drakensberg region. Colouring is proportional to a normalised stroke count, the elevation model and view used is shown in *Figure 8.1*.



(a) Negative



(b) Positive

(c) Normalised Scale for above figures

Figure 8.9: Rendered view of intracloud strokes over the Drakensberg region. Colouring is proportional to a normalised stroke count, the elevation model and view used is shown in *Figure 8.1*.

Chapter 9

Conclusion

Installation of the 19 combined TOA/MDF lightning location sensors in the Republic of South Africa began in 2005. The sensors were commissioned over the next year and a half. Subsequently 64 million strokes were detected in 801 days of data that were recorded between August 2007 and March 2010. From this data set a statistical description of lightning detected by the network has been presented.

The detection network has a well documented misclassification problem in the positive cloud-to-ground data set where intracloud strokes with small peak currents are incorrectly assigned a cloud-to-ground type classification (Wacker & Orville (1999*a*)). To enable researchers to analyse the positive cloud-to-ground data set a limit of 10.0 kA was proposed, where strokes below the limit are regarded as misclassified and discarded in an attempt to reduce the total number of misclassified strokes in the data set (Cummins *et al.* (1998)). Subsequent investigation showed that the proposed limit was too low and it was increased to 15.0 kA (Cummins & Murphy (2009)).

The statistical distribution of peak currents for the positive cloud-to-ground data set is bimodal: the first mode, the intracloud mode, is located below both the reclassification limits; and the second mode, the cloud-to-ground mode, is located above both of the reclassification limits.

The effect of moving the reclassification limit on the descriptive statistic of the positive cloud-to-ground data set is shown to dramatically reduce the number of strokes in the data set and cause a corresponding increase in the statistical measures of the distribution. Applying either of the reclassification limits effectively removes the intracloud mode, as it is located below both of the limits.

However the peak current of an intracloud stroke is not bounded below a level of cloud-to-ground activity and regardless of the magnitude of the limit applied there will always be intracloud strokes above the limit, and cloud-to-ground strokes below the limit in the positive cloud-to-ground data set.

No similar reclassification operation has been proposed for any of the other data sets, despite Biagi *et al.* (2007) failing to detect cloud-to-ground channels for small peak current strokes classified as negative cloud-to-ground in the 2003 and 2004 video study. The statistical distribution of the negative cloud-to-ground data set is unimodal, with the mode on the first reclassification limit of 10.0 kA and less than the subsequently proposed limit of 15.0 kA.

The detection of intracloud strokes by the lightning detection network is regarded as incidental, and no studies have investigated the spatial and temporal variations of intracloud data sets. The intracloud statistical distributions are unimodal, although the positive cloud-to-ground distribution has an abrupt change in gradient at the 15.0 kA reclassification limit.

Two naturally occurring probability distributions were proposed: one for the intracloud component obviously present in the positive cloud-to-ground data set and inherently present in the other data sets; and another distribution for the cloud-to-ground component also inherently present in all the data sets. It is shown that the combination of the component distributions accurately describes the statistical distribution of peak current for all of the data sets, including the abrupt gradient change in the positive intracloud distribution.

Thus there are misclassified strokes in all the data sets, but through statistical decomposition the occurrence parameters may be calculated. The network assigned type classification is not without merit through, because the dominant type in the data sets created by selecting network assigned type classification is the type selected for.

The occurrence and parameters of intracloud lightning and cloud-to-ground lightning differ based on storm parameters. The observation of positive cloud-to-ground strokes at the beginning and end of storms has been documented by Rakov & Uman (2003). However, the diurnal variation of positive cloud-to-ground lightning activity at the mean and median follows the diurnal variation of negative cloud-to-ground strokes which is contrary to field observations. The diurnal variation of the mean and median are influenced heavily by the presence of misclassified strokes in the data set and do not provide a consistent or valid measure of lightning activity.

The mode and activity at the mode provide robust statistical measures that are not sensitive the presence of outliers and to a lesser extent the measure through inspection of each mode provides a measure of the process generating the distribution of that mode. The diurnal variation of the mode of cloud-to-ground lightning stroke activity correlates with field observations.

By decomposing the positive cloud-to-ground data set into the inherently present intracloud and cloud-to-ground components the independent development of each of the components is shown. Further value is added as, in addition to a measure of activity, a measure of peak current is also obtained. Thus higher energy events present at the end of a storm are discernible against the lower energy events at the beginning of a storm.

The decomposition is also valid for the negative data sets as a prediction of the percentage of misclassified events may be determined for any of the peak current levels. These decomposition ratios compare favourably against an independent video study conducted in North America by Biagi *et al.* (2007). The modal analysis is shown to provide insight into storm activity on as fine a time scale as per day and per storm.

In addition to the diurnal variation of the statistical parameters, the orographic variation of stroke density over an area of unique severe topography is given. There orographic analysis was performed over the Drakensberg and Lesotho regions which is characterised by severe topography consisting of an escarpment of vertical walls some 500 m to 800 m high at altitudes above 3000 m a.s.l. The surrounding areas are flat with few notable orographic features.

It is shown that for the Drakensberg and Lesotho region, below 2150 m a.s.l. there is a two part relationship between altitude and stroke density and above 2150 m a.s.l. the positive cloud-to-ground data sets is invariant with altitude, and no discernible pattern is present in the negative cloud-to-ground data set. In the two part relationship initially stroke density increases with altitude and then between 1600 m a.s.l. and 1800 m a.s.l. the trend reverses and stroke density decreases with altitude. The positive intracloud data set is similar to the positive cloud-to-ground data set, but a linear relationship between stroke density and altitude is present in the negative intracloud data set.

Hence:

- the presence of misclassified strokes is shown to be present in all data sets;
- a consistent mathematical framework for the analysis of data sets consisting of multiple processes is given;
- a robust measurement of storm phase, insensitive to misclassified events, is presented;
- the diurnal trend of lightning activity over southern Africa is given; and
- the robust measurement for select days and storms is given.

9.1 Application and scope for further research

True peak current distributions, decomposed from contaminated data sets have direct application in the following areas:

- electro-geometric models for predicting transmission line shielding failure;
- risk analysis as per the IEC 62305 suite of documents; and
- meteorological now-casting to predict storm state as well as intersection time.

The following opportunities exist to extend the work further:

- the spatial distribution of strokes with specific peak currents around each sensor may yield additional information about the misclassification problem;
- there are several networks employing the same detection technology in operation around the world. The diurnal variation of the mode for select longitudes and latitudes around the world will provide insight into the global electric circuit. For detection networks active over large continental masses it will be important to confine the spatial extent or else the diurnal variation will not be discernible; and
- Lightning detection networks may also assign a probability of misclassification based on the spatial and temporal occurrence characteristics. This parameter is similar to the location error ellipse, but rather than expressing a confidence in location, the parameter expresses a confidence in type classification.

Appendix A

Statistical measures and distributions

The statistical measures and computational implementation used to characterise the probability distributions is presented in this appendix. The definitions of the mean, median, mode, moment, skewness, and kurtosis are repeated for completeness. Examples of these measures are given in terms of the widely used log-normal distribution for reference.

A.1 Measures

This section presents the measures and their implementation.

A.1.1 Probability mass function

The probability mass function is the same as a probability distribution, except it is explicitly defined for discrete random variables. The probability mass function may be calculated from a histogram (occurrence count) by normalising each bin count against the total number of samples. Hence:

$$\sum_{x} P[X = x] = 1 \tag{A.1}$$

A.1.2 Mean

The mean, or expected value, is a measure of the average response of a random variable, and is the sum of all samples divided by the number of samples. Expressed

in the continuous domain the mean is given as:

$$\bar{m} = \frac{1}{X} \int_0^X m(x)dx \tag{A.2}$$

and in a discrete form as:

$$\bar{m} = \frac{1}{N} \sum_{x=0}^{N} m[x] \tag{A.3}$$

For the log-normal distribution, the probability of the mean is defined as:

$$P_{\text{mean}} = e^{\mu + \frac{\sigma^2}{2}} \tag{A.4}$$

The expected value may be calculated from the probability mass function of a random variable as the weighed sum of the probabilities:

$$E[X] = \sum X \cdot P_x \tag{A.5}$$

where

X = Random variable

$E[X]$ = Expectation operator on random variable X

This has been implemented as follows:

Listing A.1: Octave code for `expected.m`

```
function [E] = expected(x, P)
        E = sum(x .* (P ./ sum(P)));
endfunction
```

Since the expected value of a random variable is the same as the mean of the random variable, this implementation is used to calculate the mean of the distributions.

A.1.3 Median

The median is a location measure of the point that divides the sample set in half, in other words the median is the point above which exactly half the samples are distributed and below which half the samples are distributed, for symmetrical distributions this measure is the same as the mean, but for asymmetrical distributions this measure is usually on the tail side of the distribution. In the continuous domain the median is defined as the point m, that satisfies:

$$\int_{-\infty}^{m} dF(x) \geq \frac{1}{2} \text{ and } \int_{m}^{+\infty} dF(x) \geq \frac{1}{2} \tag{A.6}$$

or in discrete form as:

$$\sum_{i=0}^{m} F(x_i) \geq \frac{1}{2} \text{ and } \sum_{i=m+1}^{n} F(x_i) \geq \frac{1}{2} \tag{A.7}$$

For the log-normal distribution, the probability of the median is defined as:

$$P_{\text{median}} = e^{\mu} \tag{A.8}$$

The median is a more robust measure of a distribution the the mean, and is insensitive to the variance of the distribution but is still influenced by the presence of extreme outliers.

A.1.4 Mode

The mode of a distribution is a measure of the most frequently occurring value. For symmetrical distributions this measure is the same as the mean and median, however for asymmetrical distributions this measure is the peak of the probability density function.

The mode provides a single concise description of a distribution and when applied to distributions originating from combined processes, the n^{th} mode can uniquely select and describe the n^{th} process. When the mode is calculated from a distribution, the measure is insensitive to outliers as it is simply the selection of the most frequently occurring result, and until the number of outliers increases substantially (i.e. well above any noise floor) the mode is invariant.

For the log-normal distribution the mode is given as:

$$P_{\text{mode}} = e^{\mu - \sigma^2} \tag{A.9}$$

For the inspected measurement of the modes in complex distributions, peaks are defined as the single point greater than the two adjacent points. This simple algorithm, computes the running differential across a section of data and takes the sign of that differential. Finally the differential of that is taken, and the point after the index where that differential is -2 corresponds to the peak.

Listing A.2: Octave code for `peak_detect.m`

```
function [X Y] = peak_detect(inX, inY)

        index = find(diff(sign(diff(inY))) == -2) +1;
        X = inX(index);
        Y = inY(index);
endfunction
```

A.1.5 Moment

The k^{th} central moment of a distribution is defined as:

$$M[X,k] = E[(X - E[X])^k] \tag{A.10}$$

where

$M[X,k]$ = Moment operator, k^{th} moment of X
k = Order of moment

And has been implemented as:

Listing A.3: Octave code for `moment.m`

```
function [mu] = moment(x, P, k)
        mu = expected((x - expected(x, P)).^k, P);
endfunction
```

A.1.6 Standard deviation

The standard deviation of a distribution is a measure of the variability or dispersion about the mean. A larger values of the standard deviation indicate that the samples are far from the mean, while smaller numbers indicate that the samples are clustered around the mean.

$$\sigma = \sqrt{E[(X - \mu)^2]} \tag{A.11}$$

A.1.7 Skewness

The skewness of a random variable is the third standardised moment and is defined as:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}} \tag{A.12}$$

A.1.8 Kurtosis

The kurtosis of a random variable is a measure of the peakedness of the probability mass function and is defined as:

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \tag{A.13}$$

The constant term in the equation above is derived by setting the kurtosis of the normal distribution to zero.

Appendix B

Algorithms and implementation

This appendix provides the implementation of the minute of the day analysis algorithm, storm phase algorithm and particle swarm optimisation algorithm.

B.1 Minute of the day analysis

The first step to processing the data is to truncate the time to minute intervals, this is done as shown in *Figure B.1.*

The code below generates the histogram by minute of the day:

Listing B.1: SQL code for `mod.sql`

```
--Create minute of day histograms, grouped by type, and current.
CREATE TABLE minute_of_day AS
(
        SELECT type,
                extract(hour FROM time)*60 +
                extract(minute FROM time) as minute_of_day,
                kA, count(kA), sum(kA)
        FROM data_distinct
        GROUP BY type, minute_of_day, kA
) WITH DATA;
```

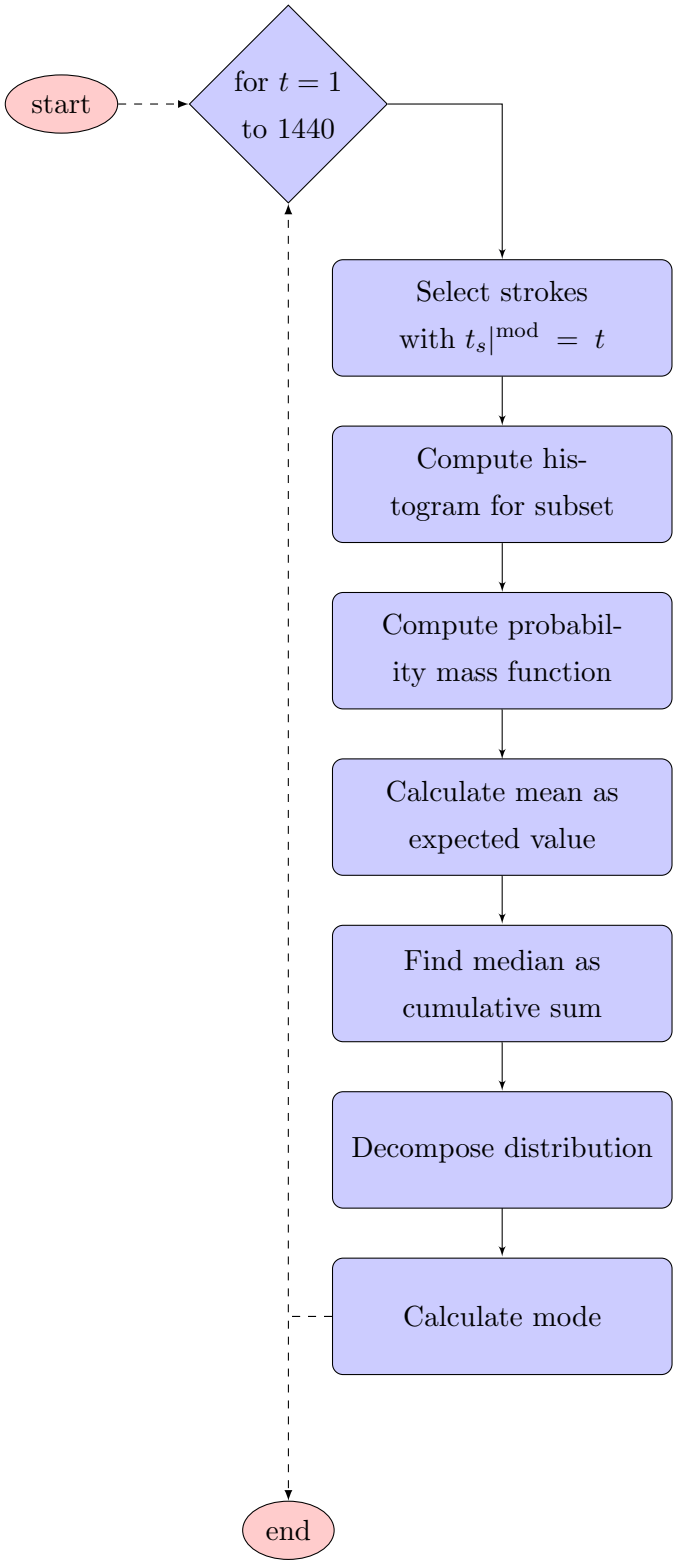


Figure B.1: Algorithm for the calculation of the temporal histogram and location of strokes, based on the minute of the day.

The code below calculates the statistical measures for each minute of the day:

Listing B.2: Octave code for `process_modhist.m`

```
modhist = dlmread('mod_hist.csv');

modhist(:,4) = modhist(:,4) / 801;  %this normalises against the
    number of days the observation was active.

ic = modhist(find(modhist(:,1) == 0), 2:4);
cg = modhist(find(modhist(:,1) == 4), 2:4);
prefix='cg';
%cg = ic; prefix = 'ic';

%figure;
peaks = zeros(1500, 4);
magnitudes = peaks;
activity = zeros(1500, 1);
means = zeros(1500, 2);
mean_mags = peaks;
medians = means;
median_mags = means;

for loop = 0:1449
        %figure;
        %subplot(2, 1, 1);
        cga = cg(find(cg(:,1) == loop), 2:3);
        pcg = cga(find(cga(:,1) >= 0), :);
        ncg = cga(find(cga(:,1) <= 0), :);
        ncg(:,1) = -1 * ncg(:,1);
        ncg = ncg(length(ncg):-1:1, :);
        [X Y] = peak_detect(pcg(:,1), pcg(:,2));
        if (size(X, 1) > 0)
                peaks(loop+1, 1) = X(1);
                magnitudes(loop+1, 1) = Y(1);
        endif

        if (size(X, 1) > 1)
                peaks(loop+1, 2) = X(2);
                magnitudes(loop+1, 2) = Y(2);
        endif

        [X Y] = peak_detect(ncg(:,1), ncg(:,2));
        if (size(X, 1) > 0)
                peaks(loop+1, 3) = X(1);
                magnitudes(loop+1, 3) = Y(1);
        endif

        if (size(X, 1) > 1)
```

```
45                  peaks(loop+1, 4) = X(2);
                    magnitudes(loop+1, 4) = Y(2);
            endif

            activity(loop+1) = sum(cga(:,2));
50          try
                    means(loop+1, 1) = expected(pcg(:,1), pcg(:,2));
                    means(loop+1, 2) = expected(ncg(:,1), ncg(:,2));
                    mean_mags(loop+1, 1) = pcg(find(pcg(:,1)>means(loop+1,
                        1))(1), 2);
                    mean_mags(loop+1, 2) = ncg(find(ncg(:,1)>means(loop+1,
                        2))(1), 2);
55          catch
            end_try_catch

            try
                    cs_neg = cumsum(ncg(:,2)./sum(ncg(:,2)));
60                  cs_pos = cumsum(pcg(:,2)./sum(pcg(:,2)));

                    pos_median_index = find(cs_pos >= 0.5)(1);
                    medians(loop+1, 1) = pcg(pos_median_index,1);
                    median_mags(loop+1, 1) = pcg(pos_median_index,2);
65
                    neg_median_index = find(cs_neg >= 0.5)(1);
                    medians(loop+1, 2) = ncg(neg_median_index,1);
                    median_mags(loop+1, 2) = ncg(neg_median_index,2);
            catch
70          end_try_catch

    endfor

    window = hanning(10);
75  window = window ./ sum(window);

    magnitudes_smooth = conv2(magnitudes, window, 'valid');
    peaks_smooth = conv2(peaks, window, 'valid');
    means_smooth = conv2(means, window, 'valid');
80  mean_mags_smooth = conv2(mean_mags, window, 'valid');
    medians_smooth = conv2(medians, window, 'valid');
    median_mags_smooth = conv2(median_mags, window, 'valid');

    dlmwrite(sprintf('%s_modes_magnitudes.dat', prefix), [(1:1440)'
        magnitudes(1:1440, :)], '\t');
85  dlmwrite(sprintf('%s_modes_peaks.dat', prefix), [(1:1440)' peaks
        (1:1440, :)], '\t');
    dlmwrite(sprintf('%s_modes_magnitudes_smooth.dat', prefix), [(1:1410)'
         magnitudes_smooth(1:1410, :)], '\t');
```

```
dlmwrite(sprintf('%s_modes_peaks_smooth.dat', prefix), [(1:1410)'
    peaks_smooth(1:1410, :)], '\t');

dlmwrite(sprintf('%s_means_peaks.dat', prefix), [(1:1440)' means
    (1:1440, :)], '\t');
dlmwrite(sprintf('%s_means_magnitudes.dat', prefix), [(1:1440)'
    mean_mags(1:1440, :)], '\t');
dlmwrite(sprintf('%s_means_peaks_smooth.dat', prefix), [(1:1410)'
    means_smooth(1:1410, :)], '\t');
dlmwrite(sprintf('%s_means_magnitudes_smooth.dat', prefix), [(1:1410)'
     mean_mags_smooth(1:1410, :)], '\t');

dlmwrite(sprintf('%s_medians_peaks.dat', prefix), [(1:1440)' medians
    (1:1440, :)], '\t');
dlmwrite(sprintf('%s_medians_magnitudes.dat', prefix), [(1:1440)'
    median_mags(1:1440, :)], '\t');
dlmwrite(sprintf('%s_medians_peaks_smooth.dat', prefix), [(1:1410)'
    medians_smooth(1:1410, :)], '\t');
dlmwrite(sprintf('%s_medians_magnitudes_smooth.dat', prefix),
    [(1:1410)' median_mags_smooth(1:1410, :)], '\t');
```

B.2 Storm phase algorithm code

This algorithm is similar to the cell tracking algorithm described by Lojou, J-Y. & Cummins (2006), but includes the ability for cells to merge. The ANSI C code below is an implementation of the algorithm presented in *Figure 7.1* on page 82.

Listing B.3: ANSI C code for `storm_track.c`

```
/****
 * Copyright (C) Michael Grant, University of the Witwatersrand 2010.
 ****/

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <unistd.h>
#include "libpq-fe.h"


#define PI 3.1415926535897932384
#define t_max 600
#define t_tick 10
#define r_initial 30
#define m_stroke 1
```

```
#define m_initial 10
#define r_growth 50
#define r_max 1000000

#define emerging_limit 20

#define GEODESIC
//#define HAVERSINE
//#define NEW_ALGORITHM

struct storm
{
  double  range[360], neighbours[360], mass[360];  //mass is the
      cumulative sum of the absolute value of peak current transferred
  double  youngest;
  double  lat, lon;
  int id;
  unsigned long int count;
  struct storm *next;
};

struct stroke
{
  double  epoch, lat, lon;
  int kA, type;
};

struct stroke_list
{
  struct stroke stroke_data;
  struct stroke_list *next;
};

int storm_counter = 0;
FILE *csv_file; // = stdout; //fopen("outsc.csv", "a");

/****
  Function clean_exit

  Closes pgsql connection and sets the return status of the program

  Arguments: Exit status and pgsql connection variable.
  Returns: void
 ****/
static void clean_exit (int exitstatus, PGconn *conn1)
{
  PQfinish(conn1);
```

```
  exit(exitstatus);
};

/****
  Function dump_storm_boundaries

  Iterates through linked list of storms, dumping extents of boundary
      in a GRASS vector form.

  Arguments: head of linked list
  Returns: void
 ****/
void dump_storm_boundaries(struct storm *head)
{
  int loop;
  struct storm *current_storm = head;

  while (current_storm != NULL)
  {
    fprintf(stdout, "L 360\n");
    //double x = 0, y = 0, mass = 0, rmax = 0;
    for (loop = 0; loop < 360; loop++)
    {
      double lat2 = asin(sin(current_storm->lat)*cos(current_storm->
          range[loop] / 6371) + cos(current_storm->lat)*sin(
          current_storm->range[loop] / 6371)*cos(PI * (double)loop /
          180));
      double lon2 = current_storm->lon + atan2(sin(PI * (double)loop /
           180)*sin(current_storm->range[loop] / 6371)*cos(
          current_storm->lat),
        cos(current_storm->range[loop] / 6371) - sin(current_storm->
           lat)*sin(lat2));

      fprintf(stdout, "%f %f\n", 180 * lon2 / PI, 180 * lat2 /PI);
    }
  }
};

/****
  Function renew_centre_of_mass

  Iterates through linked list of storms, calculating new centres
      based on the distribution of
  strokes around the previous centre of mass.

  Arguments: head of linked list
  Returns: void
```

```
 ****/
void renew_centre_of_mass (struct storm *head)
{
  int loop;
  struct storm *current_storm = head;
  double rmax = 0;
  while (current_storm != NULL)
  {
    double x = 0, y = 0, mass = 0;
    double xes[360], yes[360];

    for (loop = 0; loop < 360; loop++)
    {
      x += current_storm->mass[loop] * current_storm->range[loop] *
          cos(PI * (double)loop / 180);
      y += current_storm->mass[loop] * current_storm->range[loop] *
          sin(PI * (double)loop / 180);
      mass += current_storm->mass[loop];

      //grow the range according to the activity around the centre of
          the storm.
      current_storm->range[loop] += (current_storm->mass[loop]/mass) *
           (current_storm->neighbours[loop] - current_storm->range[
          loop]);
      if (current_storm->range[loop] < r_initial) current_storm->range
          [loop] = r_initial;
      if (current_storm->range[loop] > r_max) current_storm->range[
          loop] = r_max;
      current_storm->mass[loop] = m_initial;
      current_storm->neighbours[loop] = 0;

      xes[loop] = current_storm->range[loop] * cos(PI * (double)loop /
           180);
      yes[loop] = current_storm->range[loop] * sin(PI * (double)loop /
           180);
      //calculate points on the greate sphere so we can calculate the
          r[theta] set for the new storm centre
      //yes[loop] = asin(sin(current_storm->lat) * cos(current_storm->
          range[loop]/6371) + cos(current_storm->lat) * sin(
          current_storm->range[loop]/6371) * cos(PI * (double)loop /
          180));
      //xes[loop] =  current_storm->lon + atan2(sin(PI * (double)loop
          / 180) * sin(current_storm->range[loop]/6371) * cos(
          current_storm->lat),
      //cos(current_storm->range[loop]/6371) - sin(current_storm->lat)
          *sin(yes[loop]));
    }
```

```
        x /= mass; y /= mass;

135     double r = sqrt(pow(x, 2) + pow(y, 2));
        rmax = (r > rmax) ? r : rmax;
        double theta = atan2(sin(y)*cos(x), sin(y));  //y is latitude?
        double lat1 = asin(sin(current_storm->lat)*cos(r / 6371) + cos(
            current_storm->lat)*sin(r / 6371)*cos(theta));
        double lon1 = current_storm->lon + atan2(sin(theta)*sin(r / 6371)*
            cos(current_storm->lat), cos(r / 6371) - sin(current_storm->
            lat)*sin(lat1));
140
        //calculate new range[theta] profile
        for (loop = 0; loop < 360; loop ++)
        {
          r = sqrt(pow(xes[loop] - x, 2) + pow(yes[loop] - y, 2));
145       theta = atan2(xes[loop] - x, yes[loop] - y);


          int index = (int)round((180 / PI)*(PI + theta)); index = (index
              == 360) ? 0 : index;
          if ((index < 0) || (index > 359)) fprintf(stderr, "Index:␣\n%d\n
              ", index);
150
          current_storm->range[index] = r;
        }

        current_storm->lon = lon1; current_storm->lat = lat1;
155
        //advance to next record in linked list
        current_storm = current_storm->next;
      }
    //  fprintf(stderr, "%f !!!\n", rmax);
160 };

    /****
      Function find_nearest_storm

165   Iterates through linked list, expiring storms that are older than
          t_max (see #defines).
      Distance algorithm used is either geodesic (Haversine or Spherical
          law of cosines).
      Radial angle is snapped to nearest integral degree.

      Arguments: head of linked list, current stroke (sequentially time
          ordered).
170   Returns: pointer to closest storm, NULL if no storm found
     ****/
```

```
struct storm *find_nearest_storm (struct storm **storm_head, struct
    stroke *current_stroke, double *nextr, int *theta)
{
  double r_smallest = INFINITY;//, nextr;
  //int theta;
  struct storm *closest = NULL, *previous = NULL, *open_storms = *
      storm_head;

  //fprintf(stderr, "Starting storm search at head: %lX\n", (unsigned
      long int)*storm_head);

  while (open_storms != NULL)
  {
    if ((current_stroke->epoch - open_storms->youngest) > t_max)
    {
      //this storm has expired (no activity for too long: lets move
          along, nothing to see here folks.
      fprintf(stderr, "Expring␣storm,␣not␣enough␣activity:␣0x%lX\n", (
          unsigned long int)open_storms);
      if (open_storms == *storm_head)
      {
        fprintf(stderr, "expiring␣head:␣%lX␣->␣%lX\n", (unsigned long
            int)open_storms, (unsigned long int)open_storms->next);
        *storm_head = open_storms->next;
      }

      struct storm *next = open_storms->next;

      if (previous != NULL) previous->next = open_storms->next;
      fprintf(stderr, "Free:␣%lX\n", (unsigned long int)open_storms);
      free(open_storms);
      open_storms = next;
    }
    else
    {
      #ifdef GEODESIC
        #ifdef HAVERSINE
          *nextr = 6371 * 2 * asin(sqrt(  pow(sin((open_storms->lat -
              current_stroke->lat)/2), 2)
            + cos(open_storms->lat)*cos(current_stroke->lat)
            * pow(sin((open_storms->lon - current_stroke->lon)/2), 2))
                );
        #else
          *nextr = 6371 * acos(sin(open_storms->lat)*sin(
              current_stroke->lat)
            + cos(open_storms->lat)*cos(current_stroke->lat)
            * cos(open_storms->lon - current_stroke->lon));
```

```
210         #endif
          #else
            *nextr = pow(open_storms->lat - current_stroke->lat, 2) + pow(
                open_storms->lon - current_stroke->lon, 2);
          #endif

215         *theta = (int)round((180 / PI) * (PI + atan2(sin(current_stroke
              ->lon - open_storms->lon) * cos(current_stroke->lon),
            cos(open_storms->lat) * sin(current_stroke->lat) - sin(
                open_storms->lat) * cos(current_stroke->lat) *  cos(
                current_stroke->lon - open_storms->lon))));

          *theta = (*theta >= 360) ? (*theta - 360) : *theta;

220         if ((*theta < 0) || (*theta > 359)) fprintf(stderr, "\n%d\n", *
              theta);

          //now we see if the stroke is close enough to the storm boundary
              , we can tune this development with r_growth
          //now if we find more than one applicable storm, they must
              coalesc.
          if (*nextr < (open_storms->range[*theta] + r_growth))
225         {
            r_smallest = *nextr;
            if (closest != NULL)
            {
              //fprintf(stderr, "Coalesced %d and %d\n", open_storms->id,
                  closest->id);
230           open_storms->id = closest->id;
            }
            closest = open_storms;
          }
          previous = open_storms;
235         open_storms = previous->next;
        }
      }
      return closest;
    };
240
    /****
      Function unlink_storm

      Remove a storm from the linked list, this does NOT free the memory.
245
      Arguments: head of linked list, storm.
      Returns: void
     ****/
```

```
void unlink_storm (struct storm **storm_head, struct storm *
    unlink_storm)
{
  //fprintf(stderr, "Unlink: 0x%lX 0x%lX\n", *storm_head, unlink_storm
     );

  if (*storm_head == unlink_storm)
  {
    *storm_head = unlink_storm->next;
    return;
  }

  struct storm *current = *storm_head;
  while (current->next != unlink_storm) current = current->next;

  current->next = unlink_storm->next;
};

/****
  Function add_storm_to_list

  Add a storm structure to the end of a linked list.

  Arguments: head of linked list, new storm, and an id.
  Returns: void
 ****/
void add_storm_to_list (struct storm **storm_head, struct storm *
    emerging_storm, int storm_id)
{
  struct storm *storm_tail = *storm_head;

  while ((storm_tail != NULL) && (storm_tail->next != NULL))
    storm_tail = storm_tail->next;

  emerging_storm->id = storm_id;
  if (storm_tail == NULL)
  {
    *storm_head = emerging_storm;
    fprintf(stderr, "Adding␣head!!\n");
  }
  else storm_tail->next = emerging_storm;
  emerging_storm->next = NULL;
};

/****
  Function new_emerging_storm
```

```
  Create a new storm structure and append it to the end of the
      emerging storm list.

295   Arguments: head of linked list, current stroke (sequentially time
      ordered).
  Returns: void
 ****/
void new_emerging_storm(struct storm **emerging_storm_head, struct
    stroke *seed)
{
300   struct storm *new_storm;
  new_storm = malloc(sizeof(struct storm));
  new_storm->next = NULL;
  new_storm->lat = seed->lat;
  new_storm->lon = seed->lon;
305   new_storm->youngest = seed->epoch;
  new_storm->count = 0;

  unsigned int init_loop;
  for (init_loop = 0; init_loop < 360; init_loop++)
310   {
    new_storm->range[init_loop] = r_initial;
    new_storm->mass[init_loop] = m_initial;
  }

315   add_storm_to_list(emerging_storm_head, new_storm, 0);
};

/****
  Function add_stroke_to_list
320
  Add a stroke to the list of strokes associated with potential new
      storms.

  Arguments: list of strokes, current stroke
  Returns: void
325  ****/
void add_stroke_to_list(struct stroke_list **strokes_head, struct
    stroke current_stroke)
{
  struct stroke_list *previous = NULL, *stroke_tail = *strokes_head;

330   while ((stroke_tail != NULL) && (stroke_tail->next != NULL))
  {
    if ((current_stroke.epoch - stroke_tail->stroke_data.epoch) >
        t_max)
    {
```

```
        //this stroke has expired (no storm formed in the minimum time).
        fprintf(stderr, "Expiring stroke, no storm formed in time : 0x%
            lX\n", (unsigned long int)stroke_tail);
        if (previous != NULL) previous->next = stroke_tail->next;

        if (stroke_tail == *strokes_head) *strokes_head = stroke_tail->
            next;

        struct stroke_list *next = stroke_tail->next;
        free(stroke_tail);
        stroke_tail = next;
      }
      else
      {
        previous = stroke_tail;
        stroke_tail = stroke_tail->next;
      }
    }

    struct stroke_list *new_stroke = malloc(sizeof(struct stroke_list));

    fprintf(stderr, "New stroke: 0x%lX", (unsigned long int)new_stroke);
    new_stroke->stroke_data.epoch = current_stroke.epoch;
    new_stroke->stroke_data.lat = current_stroke.lat;
    new_stroke->stroke_data.lon = current_stroke.lon;
    new_stroke->next = NULL;

    if (stroke_tail == NULL)
    {
      fprintf(stderr, "Creating new head.\n");
      //set the list head
      *strokes_head = new_stroke;
    }
    else
    {
      fprintf(stderr, "Adding stroke to end of list\n");
      stroke_tail->next = new_stroke;
    }
  };

  /****
    Function add_stroke_to_storm

    Add a stroke to the storm, update the storm age, and mass
        distribution.

    Arguments: closest storm, current stroke
```

```
  Returns: int count of strokes in storm
 ****/
int add_stroke_to_storm(struct storm *closest, struct stroke *
    current_stroke, double nextr, int theta, int printout)
{
  closest->mass[theta] += m_stroke; //This should be set to the
      cumulative sum of the absolute value of peak current for that
      stroke; or some inverse distance thing
  closest->neighbours[theta] = (closest->neighbours[theta] > nextr) ?
      closest->neighbours[theta] : nextr;//(closest->range[theta] +
      nextr) / 2;
  closest->youngest = current_stroke->epoch;
  closest->count++;
  if (printout == 1) fprintf(csv_file, "%f,%f,%f,%d,%d,%d\n",
      current_stroke->epoch, current_stroke->lat * 180 / PI,
      current_stroke->lon  * 180 / PI, current_stroke->kA,
      current_stroke->type, closest->id);

  return closest->count;
};

void parse_stroke_list(struct stroke_list **strokes_head, struct storm
    *closest)
{
  struct stroke_list *previous = NULL, *stroke_tail = *strokes_head;
  //struct storm *find_nearest_storm (struct storm **storm_head,
      struct stroke *current_stroke, double *nextr, int *theta)
  while ((stroke_tail != NULL) && (stroke_tail->next != NULL))
  {
    double nextr; int theta;
    struct storm *found = find_nearest_storm(&closest, &(stroke_tail->
        stroke_data), &nextr, &theta);
    if (found != NULL)
    {
      add_stroke_to_storm(closest, &(stroke_tail->stroke_data), nextr,
          theta, 1);
      if (previous != NULL) previous->next = stroke_tail->next;
      free(stroke_tail);
      if (previous != NULL) stroke_tail = previous->next->next;
      else stroke_tail = NULL;
    }
    else
    {
      previous = stroke_tail;
      stroke_tail = stroke_tail->next;
    }
  }
```

```
};

/****
  Function main

  Returns: exit_status
 ****/
int main(int argc, char *argv[])
{
  const char  *conninfo = "dbname = lts";

  unsigned short int verbose = 5;

  csv_file = stdout; //fopen("outsc.csv", "a");

  char ch;
  while ((ch = getopt(argc, argv, "o:")) != -1)
  {
    switch (ch)
    {
      case 'o':
        csv_file = fopen(optarg, "w");
        break;
      case 'v':
        verbose++;
        break;
      default:
        break;
    }
  }

  PGconn    *conn;
  PGresult    *res;

  conn = PQconnectdb(conninfo);
  if (PQstatus(conn) != CONNECTION_OK)
  {
    fprintf(stderr, "Connection to database failed: %s",
        PQerrorMessage(conn));
    clean_exit(EXIT_FAILURE, conn);
  }
  else fprintf(stderr, "All ok with connections\n");

  res = PQexec(conn, "BEGIN");
  if (PQresultStatus(res) != PGRES_COMMAND_OK)
  {
    fprintf(stderr, "BEGIN failed: %s\n", PQerrorMessage(conn));
```

```
    clean_exit(EXIT_FAILURE, conn);
460 }
    else fprintf(stderr, "All␣ok␣with␣BEGIN\n");

    res = PQexec(conn, "DECLARE␣active_stroke␣CURSOR␣FOR␣SELECT␣EXTRACT(
        EPOCH␣FROM␣time),␣lat,␣lon,␣kA,␣type␣FROM␣data_distinct␣WHERE␣
        time␣BETWEEN␣'2007-11-21␣00:00:00'␣AND␣'2007-11-23␣00:00:00'␣
        ORDER␣BY␣time;");
    if (PQresultStatus(res) != PGRES_COMMAND_OK)
465 {
      fprintf(stderr, "DECLARE␣failed:␣%s\n", PQerrorMessage(conn));
      clean_exit(EXIT_FAILURE, conn);
    }
    else fprintf(stderr, "All␣ok␣with␣DECLARE\n");
470
    res = PQexec(conn, "FETCH␣100000␣IN␣active_stroke");

    struct stroke current_stroke;
    struct stroke_list *strokes_head = NULL;
475 struct storm *storm_head = NULL, *emerging_storm_head = NULL;
    int storm_id = 0;
    double tick = 0;

    while (PQresultStatus(res) == PGRES_TUPLES_OK)
480 {
      /* next, print out the rows */
      int i;

      for (i = 0; i < PQntuples(res); i++)
485   {
        if ((storm_head != NULL) && (emerging_storm_head != NULL) && (
            storm_head == emerging_storm_head))
        {
          clean_exit(EXIT_FAILURE, conn);
          fprintf(stderr, "0x%lX::0x%lX\n", (unsigned long int)
              storm_head, (unsigned long int)emerging_storm_head);
490     }

        if (!(sscanf(PQgetvalue(res, i, 0), "%lf", &current_stroke.epoch
            ) == 1) ||
          !(sscanf(PQgetvalue(res, i, 1), "%lf", &current_stroke.lat) ==
              1) ||
          !(sscanf(PQgetvalue(res, i, 2), "%lf", &current_stroke.lon) ==
              1) ||
495       !(sscanf(PQgetvalue(res, i, 3), "%d", &current_stroke.kA) ==
              1) ||
```

```
  !(sscanf(PQgetvalue(res, i, 4), "%d", &current_stroke.type) ==
      1))
{
  fprintf(stderr, "conversion␣error:␣%d\n", i);
  clean_exit(EXIT_FAILURE, conn);
}
else
{
  current_stroke.lat = current_stroke.lat * PI / 180;
  current_stroke.lon = current_stroke.lon * PI / 180;
}

if ((current_stroke.epoch - tick) > t_tick)
{
  renew_centre_of_mass(storm_head);
  tick = current_stroke.epoch;
  //if (storm_id > 10) clean_exit(EXIT_SUCCESS, conn);
}

double nextr;
int theta;
//fprintf(stderr, "Active storm: ");
struct storm *closest = find_nearest_storm(&storm_head, &
    current_stroke, &nextr, &theta);

if (closest == NULL)
{
  //no viable storm found, time to search emerging storms...
  //fprintf(stderr, "Emerging storm: ");
  closest = find_nearest_storm(&emerging_storm_head, &
      current_stroke, &nextr, &theta);
  //add_stroke_to_list(&strokes_head, current_stroke);
  if (closest == NULL)
  {
    //no emerging storm found, time to start a new one...
    if (verbose > 4) fprintf(stdout, "Creating␣new␣emerging␣
        storm.\n");
    new_emerging_storm(&emerging_storm_head, &current_stroke);
  }
  else
  {
    //emerging storm found, add to emerging storm;
    int age = add_stroke_to_storm(closest, &current_stroke,
        nextr, theta, 0);
    if (verbose > 4) fprintf(stdout, "Emerging␣storm␣found,␣
        adding␣stroke:␣%d\n", age);
    if (age >= emerging_limit)
```

```
            {
              //mature emerging storm to full storm.
              if (verbose > 4) fprintf(stdout, "Emerging storm matured.\
                  n");
              unlink_storm(&emerging_storm_head, closest);
              add_storm_to_list(&storm_head, closest, storm_counter++);
              //parse list for potential members.
              //void parse_stroke_list(struct stroke_list **strokes_head
                  , struct storm *closest)
              //parse_stroke_list(&strokes_head, closest);
            }
          }
        }
        else
        {
          //here the stroke falls within a storm boundary and so we just
               need to increase the mass at the appropriate bearing and
              rejuvinate the storm age
          add_stroke_to_storm(closest, &current_stroke, nextr, theta, 1)
              ;
        }
      }

      if (PQntuples(res) < 100000) break;
      PQclear(res);
      res = PQexec(conn, "FETCH 100000 IN active_stroke");
      fprintf(stdout, "%f,%f,%f,%d,%d,%d\n", current_stroke.epoch,
          current_stroke.lat * 180 / PI, current_stroke.lon  * 180 / PI,
           current_stroke.kA, current_stroke.type, storm_id);
    }

    if (PQresultStatus(res) != PGRES_TUPLES_OK)
    {
      fprintf(stderr, "FETCH failed: %s\n", PQerrorMessage(conn));
      clean_exit(EXIT_FAILURE, conn);
    }

    fclose(csv_file);
    clean_exit(EXIT_SUCCESS, conn);
    exit(EXIT_SUCCESS);
}
```

B.3 Particle swarm optimisation of Probability Density Function

The implementation of the algorithm described in *Figure 4.2* is given below:

Listing B.4: Octave code for `main.m`

```
%Some swarm parameters

histogram = dlmread('histogram_cg.csv');
pcg = histogram(find(histogram(:,1) >= 0), :);
ncg = histogram(find(histogram(:,1) <= 0), :);
ncg = ncg(length(ncg):-1:1, :);
ncg(:,1) = -1 * ncg(:,1);

%pcg = ncg;

pmf_f = pcg(:,2)./sum(pcg(:,2));
i2 = 2:0.1:50;
pmf = interp1(pcg(:,1), pmf_f, i2, 'spline');


SS = 100; %Number of agents
life_time = 1e6; %Number of iterations
dm = 1e-2; %Drift co-efficient (add random search space)
dmr = 2*dm;
dg = 0.1; %Global fitness drift
pm = 2; %particle mass.

%R1 = 1e5*rand(1, SS);
%R2 = 1e5*ones(1, SS);
%C1 = 1e-20*ones(1, SS);
%C2 = 1e-20*ones(1, SS);
%k = sqrt(15);%k = 10*rand(1, SS);  %Search space, 5 dimensions

A1 = 0*rand(1, SS); %try 1 here.
A2 = 10*rand(1, SS); %try 1 here.

%logrnd paramters:
MU1 = 3 + 0.5*rand(1, SS); SIGMA1 = 3*rand(1, SS); dm=1e-1; pm=2; dmr
    =10*dm;
MU2 = 2 + 0.5*rand(1, SS); SIGMA2 = 1*rand(1, SS);

%gampdf paramters:
%MU2 = 10*rand(1, SS); SIGMA2 = 1*rand(1, SS);  %These parameters work
     well with @gampdf (try 1 with MU2;
```

```
%Cauchy_pdf parameters:
%MU1 = 20*rand(1, SS); SIGMA1 = 5*rand(1, SS);  %These parameters work well with @cauchy_pdf
%MU2 = 20*rand(1, SS); SIGMA2 = 5*rand(1, SS);  %These parameters work well with @cauchy_pdf

%T1 = 1.5*rand(1, SS); T2 = 1.5*rand(1, SS);
T1 = ones(1, SS); T2 = ones(1,SS);

VA1 = zeros(1, SS); VA2 = VA1; VMU1 = VA1; VMU2 = VA1; VS1 = VA1; VS2 = VA1; VT1 = VA1; VT2 = VA1;


gcosta = inf;
dispersion = inf;
fc = 1;

dist1 = @lognpdf;%@cauchy_pdf;
dist2 = @lognpdf;%@gampdf;

nochange = 0;
count = 0;
do
        %Calculate swarm member cost function.
        %a = R1.*R2.*C1.*C2;
        %b = R1.*C1 + R2.*C1 + R1.*C2.*(1-k);
        %c = 1;

        %zfa = (-b + sqrt(b.^2 - 4.*a.*c))./(2*a);
        %zfb = (-b - sqrt(b.^2 - 4.*a.*c))./(2*a);

        %costa = abs(zp - zfa);
        %costb = abs(zp - zfb);

        for loop = 1:SS
                a = A1(loop)*dist1(T1(loop) * i2, MU1(loop), SIGMA1(loop));
                b = A2(loop)*dist2(T2(loop) * i2, MU2(loop), SIGMA2(loop));
                costa(loop) = sum(abs(pmf - (a + b)));
        endfor

        if (min(costa) < gcosta)
                fc = find(costa == min(costa));
                fc = fc(1);
                %fR1 = R1(fc); fR2 = R2(fc); fC1 = C1(fc); fC2 = C2(fc); %fk = k(fc);
```

```
80                  fA1 = A1(fc); fA2 = A2(fc);
                    fMU1 = MU1(fc); fMU2 = MU2(fc);
                    fSIGMA1 = SIGMA1(fc); fSIGMA2 = SIGMA2(fc);
                    fT1 = T1(fc); fT2 = T2(fc);
                    gcosta = min(costa);
85                  fprintf(stderr, '%f %d\t', gcosta, nochange);
                    fprintf(stderr, '%f %f %f %f - %f %f %f %f\n', fA1,
                        fMU1, fSIGMA1, fT1, fA2, fMU2, fSIGMA2, fT2);
                    nochange = 0;
                    %figure(1);
                    %plot(R1, R2, 'bx', fR1, fR2, 'rx');
90                  %figure(2);
                    %plot(C1, C2, 'x', fC1, fC2, 'rx');
                    %    plot(costa, R1, 'x', gcosta, fR1, 'rx');
                    figure(1);
                    plot3([A1 A2], [MU1 MU2], [SIGMA1 SIGMA2], 'x');
95                  figure(2);
                    a = fA1*dist1(fT1* i2, fMU1, fSIGMA1);
                    b = fA2*dist2(fT2* i2, fMU2, fSIGMA2);
                    plot(i2, [pmf; a + b], pcg(1:25,1), pmf_f(1:25), 'r-x
                        ');
                    figure(3);
100                 plot(i2, [a; b]);
                    pause(1e-1);
                    dispersion = max(costa) - min(costa);
                    fprintf(stderr, "Dispersion: %f\n", dispersion);
            elseif
105                 %fprintf(stderr, 'o');
                    nochange++;
            endif

            %Find the vector to the fittest swarm member
110         %dR1 = fR1 - R1; dR2 = fR2 - R2; dC1 = fC1 - C1; dC2 = fC2 -
                C2; %dk = fk - k;
            dA1 = fA1 - A1; dA2 = fA2 - A2;
            dMU1 = fMU1 - MU1; dMU2 = fMU2 - MU2;
            dSIGMA1 = fSIGMA1 - SIGMA1; dSIGMA2 = fSIGMA2 - SIGMA2;
            dT1 = fT1 - T1; dT2 = fT2 - T2;
115
            %add drift, less when there is a better fit (aka preserve
                global maximum)
            costa = 1/(10^(nochange/20));
            %dR1 = dg*dR1 + dmr*costa.*(rand(1, SS) - 0.5);
            %dR2 = dg*dR2 + dmr*costa.*(rand(1, SS) - 0.5);
120         %dC1 = dg*dC1 + dm*costa.*(rand(1, SS) - 0.5);
            %dC2 = dg*dC2 + dm*costa.*(rand(1, SS) - 0.5);
            %dk = dg*dk + dm*costa.*(rand(1, SS) - 0.5);
```

```
            %Calculate the applied force:
            %dA1 = dg*dA1 + dm*costa.*(rand(1, SS) - 0.5);
            %dA2 = dg*dA2 + dm*costa.*(rand(1, SS) - 0.5);
            %dMU1 = dg*dMU1 + dmr*costa.*(rand(1, SS) - 0.5);
            %dMU2 = dg*dMU2 + dmr*costa.*(rand(1, SS) - 0.5);
            %dSIGMA1 = dg*dSIGMA1 + dm*costa.*(rand(1, SS) - 0.5);
            %dSIGMA2 = dg*dSIGMA2 + dm*costa.*(rand(1, SS) - 0.5);
            %dT1 = dg*dT1 + dm*costa.*(rand(1, SS) - 0.5);
            %dT2 = dg*dT2 + dm*costa.*(rand(1, SS) - 0.5);
            dA1 = dg*dA1 + dm*costa.*randn(1, SS);
            dA2 = dg*dA2 + dm*costa.*randn(1, SS);
            dMU1 = dg*dMU1 + dmr*costa.*randn(1, SS);
            dMU2 = dg*dMU2 + dmr*costa.*randn(1, SS);
            dSIGMA1 = dg*dSIGMA1 + dm*costa.*randn(1, SS);
            dSIGMA2 = dg*dSIGMA2 + dm*costa.*randn(1, SS);
            dT1 = dg*dT1 + dm*costa.*randn(1, SS);
            dT2 = dg*dT2 + dm*costa.*randn(1, SS);

            VA1 = VA1/pm + dA1; VA2 = VA2/pm + dA2;
            VMU1 = VMU1/pm + dMU1; VMU2 = VMU2/pm + dMU2;
            VS1 = VS1/pm + dSIGMA1; VS2 = VS2/pm + dSIGMA2;
            VT1 = VT1/pm + dT1; VT2 = VT2/pm + dT2;

            %Move the particles in the swarm
            %R1 = R1 + dR1; R2 = R2 + dR2; C1 = C1 + dC1; C2 = C2 + dC2; %
                k = k + dk;
            %A1 = A1 + dA1; A2 = A2 + dA2; MU1 = MU1 + dMU1; MU2 = MU2 +
                dMU2; SIGMA1 = SIGMA1 + dSIGMA1; SIGMA2 = SIGMA2 + dSIGMA2
                ;
            A1 = A1 + VA1; A2 = A2 + VA2;
            MU1 = MU1 + VMU1; MU2 = MU2 + VMU2;
            SIGMA1 = SIGMA1 + VS1; SIGMA2 = SIGMA2 + VS2;
            %T1 = T1 + VT1; T2 = T2 + VT2;

            %Now confine ourselves to real valued vectors...
            %R1 = abs(R1); R2 = abs(R2); C1 = abs(C1); C2 = abs(C2);
            A1 = abs(A1); A2 = abs(A2);
            MU1 = abs(MU1); MU2 = abs(MU2);
            SIGMA1 = abs(SIGMA1); SIGMA2 = abs(SIGMA2);
            T1 = abs(T1); T2 = abs(T2);
            count++;
    until ((nochange > 200) || ((dispersion < 0.01) && (dispersion != 0)))
        ;

    fprintf(stderr, '\n');
```

References

Betz, H., Schumann, U. & Laroche, P., eds (2008), *Lightning: Principles, Instruments and Applications*, first edn, Springer Science. ISBN 978-1-4020-9078-3.

Biagi, C., Cummins, K., Kehoe, K. & Krider, E. (2007), 'National lightning detection network (NLDN) performance in southern Arizona, Texas and Oklahoma in 2003–2004', *Journal of Geophysical Research*, vol. 112, no. D05208, pp. 1–17.

Boccippio, D., Cummins, K., Christian, H. & Goodman, S. (2001), 'Combined satellite- and surface-based estimation of the intracloud-cloud-to-ground lightning ratio over the continental united states', *Monthly Weather Review*, vol. 129, pp. 108–122.

Chisholm, W. & Cummins, K. (2006), Lightning parameters: A review, applications and extensions, *in* 'Proceedings of IEEE T&D', Panel PN-11, May 2006, Dallas, Texas, United States of America.

Chiu, C.-S. & Klett, J. (1976), 'Convective electrification of clouds', *Journal of Geophysical Research*, vol. 81, no. 6, pp. 1111–1124.

Chowdhuri, P., Anderson, J., Chisholm, W., Field, T., Ishii, M., Martinez, J., Marz, M., J., M., McDermott, T., Mousa, A., Narita, T., Nichols, D. & Short, T. (2005), 'Parameters of lightning strokes: A review', *IEEE Transactions on Power Delivery*, vol. 20, no. 1, pp. 346–358.

Christian, H., Blakeslee, R., Boccippio, D., Boeck, W., Buechler, D., Driscoll, K., Goodman, S., Hall, J., Koshak, W., Mach, D. & Stewart, M. (2003), 'Global frequency and distribution of lightning as observed from space by the optical transient detector', *Journal of Geophysical Research*, vol. 108, no. D1,4005, pp. 1–15.

Cooray, G. (2003), *The Lightning Flash*, 1st edn, The Institution of Electrical Engineers, Michael Faraday House, Six Hills Ways, Stevenage, SG1 2AY, United Kingdom. ISBN 0 85296 780 2.

Cummins, K. & Murphy, M. (2009), 'An overview of lightning locating systems: History, techniques and data uses, with an in-depth look at the U.S. NLDN', *IEEE Transactions on Electromagnetic Compatibility*, vol. 51, no. 3, pp. 499–518.

Cummins, K., Cramer, J., Biagi, C., Krider, E., Jerauld, J., Uman, M. & Rakov, V. (2006), The U.S. National Lightning Detection Network: Post-upgrade status, *in* 'Second Conference on Meteorological Applications of Lightning Data', American Meteorological Society, Atlanta, Georgia.

Cummins, K., Murphy, M., Bardo, E., Hiscox, W.L. Pyle, R. & Pifer, A. (1998), 'A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network', *Journal of Geophysical Research*, vol. 103, no. D8, pp. 9035–9044.

Donner, L. & Phillips, V. (2003), 'Boundary layer control on convective available potential energy: Implications for cumulus parameterization', *Journal of Geophysical Research*, vol. 108, no. D22, pp. 4701–4712.

Franklin, B. (1753), *Papers of Benjamin Franklin*, Vol. 1–31, Yale University Press, New Haven, Connecticut, United States of America. Poor Richards Almanac for 1753.

Gill, T. (2008), Initial steps in the development of a comprehensive lightning climatology of south africa, Master's dissertation, School of Geography, University of the Witwatersrand, Johannesburg, South Africa.

Grohmann, C. (2004), 'Morphometric analysis in geographic information systems: applications of free software GRASS and R*', *Computers and Geosciences*, vol. 30, pp. 1055–1067.

Idone, V., Saljoughy, A., Henderson, R., Moore, P. & Pyle, R. (1993), 'A re-examination of the peak current calibration of the national lightning detection network', *Journal of Geophysical Research*, vol. 98, no. D10, pp. 18323–18332.

Jennings, S. (1975), 'Charge separation due to water drop and cloud droplet interactions in an electric field', *Quarterly Journal of the Royal Meteorological Society*, vol. 101, no. 428, pp. 227–233.

Jerauld, J., Rakov, V., Uman, M., Rambo, K. & Jordan, D. (2005), 'An evaluation of the performance characteristics of the U.S. National Lightning Detection Network in Florida using rocket-triggered lightning', *Journal of Geophysical Research*, vol. 110, no. D19106, pp. 1–16.

Koshak, W. & Solakiewicz, R. (2001), 'TOA lightning location retrieval on spherical and oblate spheroidal earth geometries', *Journal of Atmospheric and Oceanic Technology*, vol. 18, pp. 187–199.

Krider, E., Noggle, R. & Uman, M. (1976), 'A gated wide band magnetic direction finder for lightning return strokes', *Journal of Applied Meteorology*, vol. 15, pp. 301–306.

Lang, T. & Rutledge, S. (2002), 'Relationships between convective storm kinematics, precipitation and lightning', *Monthly Weather Review*, vol. 130, pp. 2492–2506.

Lojou, J-Y. & Cummins, K. (2006), Total lightning mapping using both VHF interferometry and time-of-arrival techniques, *in* '19$^{\text{th}}$ International Lightning Detection conference and 1$^{\text{st}}$ International Lightning Meteorology Conference', Tucson, Arizona, United States of America, pp. 1–9.

Mazur, V. (2002), 'Physical processes during development of lightning flashes', *C. R. Physique*, vol. 3, no. 10, pp. 1393–1409.

Oppenheim, A., Schafer, R. & Buck, J. (1998), *Discrete-time signal processing*, 1 edn, Prentice-Hall, Upper Saddle River, New Jersey 07458.

Price, C., Melnikov & A. (2004), 'Diurnal, seasonal and inter-annual variations in the Schumann resonance parameters', *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 66, pp. 1179–1185.

Rakov, V. & Uman, M. (2003), *Lightning Physics and Effects*, 1$^{\text{st}}$ edn, Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, United Kingdom. ISBN 0 521 58327 6.

Schonland, B. (1956), 'The lightning discharge', *Handbuch der Physik*, vol. 22, pp. 576–628.

Thompson, E., Medelius, P. & Davis, S. (1994), 'System for locating the sources of wide band de/dt from lightning', *Journal of Geophysical Research*, vol. 99, no. D11, pp. 22793–22802.

Thottappillil, R., Rakov, V. & Uman, M. (1997), 'Distribution of charge along the lightning channel: relation to remote electric and magnetic fields and to return stroke models', *Journal of Geophysical Research*, vol. 102, pp. 6987–7006.

Thottappillil, R., Uman, M. & Rakov, V. (1998), 'Treatment of retardation effects in calculating the radiated electromagnetic fields from the lightning discharge', *Journal of Geophysical Research*, no. 103, pp. 9003–9013.

Tsenova, B., Mitzeva, R. & Saunders, C. (2009), 'A modelling study of the effect of ice particle sized and relative velocity on ice crystal/graupel collisional charge transfer', *Atmospheric Research*, vol. 91, no. 2–4, pp. 250–258.

Wacker, R. & Orville, R. (1999*a*), 'Changes in measured lightning flash count and return stroke peak current after the 1994 U.S. National Lightning Detection Network Upgrade: 1. Observations', *Journal of Geophysical Research*, vol. 104, no. D2, pp. 2151–2157.

Wacker, R. & Orville, R. (1999*b*), 'Changes in measured lightning flash count and return stroke peak current after the 1994 U.S. National Lightning Detection Network Upgrade: 2. Theory', *Journal of Geophysical Research*, vol. 104, no. D2, pp. 2159–2162.

Zajac, B. & Rutledge, S. (2001), 'Cloud-to-ground lightning activity in the contiguous United States from 1995 to 1999', *Monthly Weather Review*, vol. 129, pp. 999–1019.

Bibliography

Anderson, R. (1971), A comparison between some lightning parameters measured in Switzerland with those in Southern Africa, Technical report, CIGRE Study Committee No. 33.01.

Betz, H., Schumann, U. & Laroche, P., eds (2008), *Lightning: Principles, Instruments and Applications*, first edn, Springer Science. ISBN 978-1-4020-9078-3.

Biagi, C., Cummins, K., Kehoe, K. & Krider, E. (2007), 'National lightning detection network (NLDN) performance in southern Arizona, Texas and Oklahoma in 2003–2004', *Journal of Geophysical Research*, vol. 112, no. D05208, pp. 1–17.

Boccippio, D., Cummins, K., Christian, H. & Goodman, S. (2001), 'Combined satellite- and surface-based estimation of the intracloud-cloud-to-ground lightning ratio over the continental united states', *Monthly Weather Review*, vol. 129, pp. 108–122.

Bowdler, G. (1973), *Measurements in High-voltage Test Circuits*, first edn, Pregamon Press Ltd, Maxwell House, Fairview Park, Elmsford, New York, 10523.

Chisholm, W. & Cummins, K. (2006), Lightning parameters: A review, applications and extensions, *in* 'Proceedings of IEEE T&D', Panel PN-11, May 2006, Dallas, Texas, United States of America.

Chiu, C.-S. & Klett, J. (1976), 'Convective electrification of clouds', *Journal of Geophysical Research*, vol. 81, no. 6, pp. 1111–1124.

Chowdhuri, P., Anderson, J., Chisholm, W., Field, T., Ishii, M., Martinez, J., Marz, M., J., M., McDermott, T., Mousa, A., Narita, T., Nichols, D. & Short, T. (2005), 'Parameters of lightning strokes: A review', *IEEE Transactions on Power Delivery*, vol. 20, no. 1, pp. 346–358.

Christian, H., Blakeslee, R., Boccippio, D., Boeck, W., Buechler, D., Driscoll, K., Goodman, S., Hall, J., Koshak, W., Mach, D. & Stewart, M. (2003), 'Global frequency and distribution of lightning as observed from space by the optical

transient detector', *Journal of Geophysical Research*, vol. 108, no. D1,4005, pp. 1–15.

Cooray, G. (2003), *The Lightning Flash*, 1st edn, The Institution of Electrical Engineers, Michael Faraday House, Six Hills Ways, Stevenage, SG1 2AY, United Kingdom. ISBN 0 85296 780 2.

Cooray, V., Becerra, M. & Rakov, V. (2006), On the electric field at the tip of dart leaders in lightning flashes, *in* 'Proceedings of the 28^{th} International Conference on Lightning Protection', Kanazawa, Japan.

Cooray, V., Fernando, M., Sörensen, T., Götschl, T. & Pedersen, A. (2000), 'Propagation of lightning generated transient electromagnetic fields over finitely conducting ground', *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 62, pp. 583–600.

Cooray, V., Zitnik, M. & Manyahi, M. (2004), 'Physical model of surge-current characteristics of buried vertical rods in the presence of soil ionisation', *Journal of Electrostatics*, vol. 60, pp. 193–202.

Cummins, K. & Murphy, M. (2009), 'An overview of lightning locating systems: History, techniques and data uses, with an in-depth look at the U.S. NLDN', *IEEE Transactions on Electromagnetic Compatibility*, vol. 51, no. 3, pp. 499–518.

Cummins, K., Cramer, J., Biagi, C., Krider, E., Jerauld, J., Uman, M. & Rakov, V. (2006), The U.S. National Lightning Detection Network: Post-upgrade status, *in* 'Second Conference on Meteorological Applications of Lightning Data', American Meteorological Society, Atlanta, Georgia.

Cummins, K., Murphy, M., Bardo, E., Hiscox, W.L. Pyle, R. & Pifer, A. (1998), 'A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network', *Journal of Geophysical Research*, vol. 103, no. D8, pp. 9035–9044.

Donner, L. & Phillips, V. (2003), 'Boundary layer control on convective available potential energy: Implications for cumulus parameterization', *Journal of Geophysical Research*, vol. 108, no. D22, pp. 4701–4712.

Dwyer, J. (2004), 'Implications of x-ray emission from lightning', *Geophysical Research Letters*, vol. 31, pp. L12102.

Dwyer, J. (2005), 'The initiation of lightning by runaway air breakdown', *Geophysical Research Letters*, vol. 32, pp. L20808.

Dwyer, J., Rassoul, H., Al-Dayeh, M., Caraway, L., Wright, B., Chrest, A., Uman, M., Rakov, V., Rambo, K., Jordan, D., Jerauld, J. & Smyth, C. (2004), 'Measurements of x-ray emission from rocket-triggered lightning', *Geophysical Research Letters*, vol. 31, pp. L05118.

Dwyer, J., Rassoul, H., Saleh, Z., Uman, M., Jerauld, J. & Plumer, J. (2005), 'X-ray bursts produced by laboratory sparks in air', *Geophysical Research Letters*, vol. 32, pp. L20809.

Forest, J. L. (1982), *Transmission Line Reference book*, second edn, Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California.

Franklin, B. (1753), *Papers of Benjamin Franklin*, Vol. 1–31, Yale University Press, New Haven, Connecticut, United States of America. Poor Richards Almanac for 1753.

Gallagher, T. & Pearmain, A. (1983), *High Voltage Measurement, Testing and Design*, John Wiley and Sons.

Gallimberti, I., Bacchiega, G., Bondiou-Clergerie, A. & Lalande, P. (2002), 'Fundamental processes in long air gap discharges', *C. R. Physique*, vol. 3, pp. 1335–1359.

Gao, Y., He, J., Zou, J., Zeng, R. & Liang, X. (2005), 'Fractal simulation of soil breakdown under lightning current', *Journal of Electrostatics*, vol. 61, pp. 197–207.

Gill, T. (2008), Initial steps in the development of a comprehensive lightning climatology of south africa, Master's dissertation, School of Geography, University of the Witwatersrand, Johannesburg, South Africa.

Goto, Y. & Narita, K. (2006), Summary of winter lightning at maki in the japan sea coastal area, *in* 'Proceedings of the 28^{th} International Conference on Lightning Protection', Kanazawa, Japan.

Grohmann, C. (2004), 'Morphometric analysis in geographic information systems: applications of free software GRASS and R*', *Computers and Geosciences*, vol. 30, pp. 1055–1067.

Gurevich, A., Milikh, G. & Roussel-Dupre, R. (1999*a*), 'Runaway electron mechanism of air breakdown and preconditioning during a thunderstorm', *Physics Letters A*, vol. 254, pp. 79–87.

Gurevich, A., Zybin, K. & Roussel-Dupre, R. (1999*b*), 'Lightning initiation by simultaneous effect of runaway breakdown and cosmic ray showers', *Physics Letters A*, vol. 254, pp. 79–87.

Hamlin, T., Light, T. E., Shao, X. M., Eack, K. B. & Harlin, J. D. (2007), 'Estimating lightning channel characteristics of positive narrow bipolar events using intrachannel current reflection signatures', *Journal of Geophysical Research*, vol. 112, pp. D14108.

Heidler, F., Cvetic, J. & Stanic, B. (1999), 'Calculation of lightning current parameters', *IEEE Transactions on Power Delivery*, vol. 14, no. 2, pp. 399–404.

Horváth, T. (2006), Interception efficiency of lightning air termination systems constructed with rolling sphere method, *in* 'Proceedings of the 28^{th} International Conference on Lightning Protection, Kanazawa, Japan. 2006', Kanazawa, Japan.

Hu, W., Cummer, S. A. & Lyons, W. A. (2007), 'Testing sprite initiation theory using lightning measurements and modeled electromagnetic fields', *Journal of Geophysical Research*, vol. 112, pp. D13115.

Idone, V., Saljoughy, A., Henderson, R., Moore, P. & Pyle, R. (1993), 'A reexamination of the peak current calibration of the national lightning detection network', *Journal of Geophysical Research*, vol. 98, no. D10, pp. 18323–18332.

Jackson, J. (1999), *Classical Electrodynamics*, first edn, John Wiley and Sons, 605 Third Avenue, New York, USA.

James, G., Burley, D., Clements, D., Dyke, P., Searl, J., Steele, N. & Wright, J. (1999), *Advanced Modern Engineering Mathematics*, second edn, Pearson Education Limited, Edinburgh Gate, Harlow, Essex, CM20 2JE, England.

Jennings, S. (1975), 'Charge separation due to water drop and cloud droplet interactions in an electric field', *Quarterly Journal of the Royal Meteorological Society*, vol. 101, no. 428, pp. 227–233.

Jerauld, J., Rakov, V., Uman, M., Rambo, K. & Jordan, D. (2005), 'An evaluation of the performance characteristics of the U.S. National Lightning Detection Network in Florida using rocket-triggered lightning', *Journal of Geophysical Research*, vol. 110, no. D19106, pp. 1–16.

Jerauld, J., Uman, M. A., Rakov, V. A., Rambo, K. J. & Schnetzer, G. H. (2007), 'Insights into the ground attachment process of natural lightning gained from an unusual triggered-lightning stroke', *Journal of Geophysical Research*, vol. 112, pp. D13113.

Kordi, B., Moini, R., Janischewskyj, W., Hussein, A., Shostak, V. & Rakov, V. (2003), 'Application of antenna theory model to a tall tower struck by lightning', *Journal of Geophysical Research*, vol. 108, no. D17, pp. 7.1–7.9.

Koshak, W. & Solakiewicz, R. (2001), 'TOA lightning location retrieval on spherical and oblate spheroidal earth geometries', *Journal of Atmospheric and Oceanic Technology*, vol. 18, pp. 187–199.

Kosztaluk, R., Loboda, M. & Mukhedkar, D. (1981), 'Experimental study of transient ground impedances', *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 11, pp. 4653–4659.

Krider, E., Noggle, R. & Uman, M. (1976), 'A gated wide band magnetic direction finder for lightning return strokes', *Journal of Applied Meteorology*, vol. 15, pp. 301–306.

Kuffel, E., Zaengl, W. & Kuffel, J. (2000), *High Voltage Engineering: Fundamentals*, second edn, Newnes, Linacre House, Jordan Hill, Oxford OX2 8DP, England.

Lalande, P., Bondiou-Clergerie, A., Bacchiega, G. & Gallimberti, I. (2002), 'Observations and modeling of lightning leaders', *C. R. Physique*, vol. 3, pp. 1375–1392.

Lang, T. & Rutledge, S. (2002), 'Relationships between convective storm kinematics, precipitation and lightning', *Monthly Weather Review*, vol. 130, pp. 2492–2506.

Lojou, J-Y. & Cummins, K. (2006), Total lightning mapping using both VHF interferometry and time-of-arrival techniques, *in* '19$^{\text{th}}$ International Lightning Detection conference and 1$^{\text{st}}$ International Lightning Meteorology Conference', Tucson, Arizona, United States of America, pp. 1–9.

Mazur, V. (2002), 'Physical processes during development of lightning flashes', *C. R. Physique*, vol. 3, no. 10, pp. 1393–1409.

Mazzetti, C. & Veca, G. (1983), 'Impulse behaviour of ground electrodes', *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 9, pp. 3148–3156.

Milikh, G. & Valdivia, J. (1999), 'Model of gamma ray flashes due to fractal lightning', *Geophysical Research Letters*, vol. 26, no. 4, pp. 525–528.

Montanya, J., Soula, S. & Pineda, N. (2007), 'A study of the total lightning activity in two hailstorms', *Journal of Geophysical Research*, vol. 112, pp. D13118.

Oppenheim, A., Schafer, R. & Buck, J. (1998), *Discrete-time signal processing*, 1 edn, Prentice-Hall, Upper Saddle River, New Jersey 07458.

Orville, R., Huffines, G., Burrows, W., Holle, R. & Cummins, K. (2002), 'The North American Lightning Detection Network (NLDN) – First results: 1998-2000', *Monthly Weather Review*, vol. 130, pp. 2098–2108.

Pavanello, D. & et al. (2007), 'On return stroke currents and remote electromagnetic fields associated with lightning strikes to tall structures: 2. experiment and model validation', *Journal of Geophysical Research*, vol. 112, pp. D13122.

Phillips, A., Grobbelaar, G. & Jandrell, I. (1999*a*), The modelling and development of a resistively wound Rogowski coil for measuring lightning current impulses, *in* 'South African Universities Power Engineering conference'. Electric Power Research Group, University of the Witwatersrand.

Phillips, A., Grobbelaar, G., Pritchard, C., Melaia, R. & Jandrell, I. (1999*b*), 'Development of a Rogowski coil to measure lightning current impulses', *Transactions of the SAIEE*.

Pigini, A., Rizzi, G., Garbagnati, E., Porrino, A., Baldo, G. & Pesavento, G. (1989), 'Performance of large air gaps under lightning over voltages: Experimental study and analysis of accuracy of predetermination methods', *IEEE Transactions on Power Delivery*, vol. 4, no. 2, pp. 1379–1392.

Price, C., Melnikov & A. (2004), 'Diurnal, seasonal and inter-annual variations in the Schumann resonance parameters', *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 66, pp. 1179–1185.

Rakov, V. & Uman, M. (2003), *Lightning Physics and Effects*, 1st edn, Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, United Kingdom. ISBN 0 521 58327 6.

Rakov, V., Uman, M. & Rambo, K. (2005), 'A review of ten years of triggered-lightning experiments at Camp Blanding, Florida', *Journal of Atmospheric Research*, vol. 76, pp. 503–517.

Riousset, J. A., Pasko, V. P., Krehbeil, P. R., Thomas, R. J. & Rison, W. (2007), 'Three-dimensional fractal modeling of intracloud lightning discharge in a new mexico thunderstorm and comparison with lightning mapping observations', *Journal of Geophysical Research*, vol. 112, pp. D15203.

Schonland, B. (1950), *The flight of thunderbolts*, Oxford University Press, Amen house, London, England.

Schonland, B. (1956), 'The lightning discharge', *Handbuch der Physik*, vol. 22, pp. 576–628.

Serway, R. A., Beichner, R. J. & Jewett, J. W. (2000), *Physics for Scientists and Engineers*, fifth edn, Saunders College Publishing, 6277 Sea Harbour drive, Orlando, FL 32887-6777, USA.

Solomon, R., Adamo, C. & Baker, M. (2002), 'A lightning initiation mechanism: application to a thunderstorm electrification model', *C. R. Physique*, vol. 3, pp. 1325–1333.

Soriano, L. R. & de Pablo, F. (2007), 'Total flash density and the intracloud/cloud-to-ground lightning ratio over the iberian peninsula', *Journal of Geophysical Research*, vol. 112, pp. D13114.

Stanley, M., Shao, X., Smith, D., Lopez, L., Pongratz, M., Harlin, J., Stock, M. & Regan, A. (2006), 'A link between terrestrial gamma-ray flashes and intracloud lightning discharges', *Geophysical Research Letters*, vol. 33, pp. L06803.

Thompson, E., Medelius, P. & Davis, S. (1994), 'System for locating the sources of wide band de/dt from lightning', *Journal of Geophysical Research*, vol. 99, no. D11, pp. 22793–22802.

Thottappillil, R., Rakov, V. & Uman, M. (1997), 'Distribution of charge along the lightning channel: relation to remote electric and magnetic fields and to return stroke models', *Journal of Geophysical Research*, vol. 102, pp. 6987–7006.

Thottappillil, R., Rakov, V. A. & Theethayi, N. (2007), 'Expressions for far electric fields produces at an arbitrary altitude by lightning return strokes', *Journal of Geophysical Research*, vol. 112, pp. D16102.

Thottappillil, R., Uman, M. & Rakov, V. (1998), 'Treatment of retardation effects in calculating the radiated electromagnetic fields from the lightning discharge', *Journal of Geophysical Research*, no. 103, pp. 9003–9013.

Tsenova, B., Mitzeva, R. & Saunders, C. (2009), 'A modelling study of the effect of ice particle sized and relative velocity on ice crystal/graupel collisional charge transfer', *Atmospheric Research*, vol. 91, no. 2–4, pp. 250–258.

Vargas, M., Cooray, V., Becerra, M. & Torres, H. (2006), Lightning leader channel modelling, *in* 'Proceedings of the 28^{th} International Conference on Lightning Protection', Kanazawa, Japan.

Wacker, R. & Orville, R. (1999a), 'Changes in measured lightning flash count and return stroke peak current after the 1994 U.S. National Lightning Detection Network Upgrade: 1. Observations', *Journal of Geophysical Research*, vol. 104, no. D2, pp. 2151–2157.

Wacker, R. & Orville, R. (1999*b*), 'Changes in measured lightning flash count and return stroke peak current after the 1994 U.S. National Lightning Detection Network Upgrade: 2. Theory', *Journal of Geophysical Research*, vol. 104, no. D2, pp. 2159–2162.

Ward, D. & Exon, J. (1993), 'Using rogowski coils for transient current measurements', *Engineering Science and Education Journal*, pp. 105–113.

Williams, E., Weber, M. & Orville, R. (1989), 'The relationship between lightning type and convective state of thunderclouds', *Journal of Geophysical Research*, vol. 94, no. D11, pp. 13213–13220.

Wilson, B. & Ghassemlooy, Z. (1985), 'Optical pulse width modulation for electrically isolated analogue transmission', *Journal of Physics and Electrical Scientific Instrumentation*, vol. 15, pp. 954–958.

Zajac, B. & Rutledge, S. (2001), 'Cloud-to-ground lightning activity in the contiguous United States from 1995 to 1999', *Monthly Weather Review*, vol. 129, pp. 999–1019.